THE EFFECTS OF TASK COMPLEXITY, MODE OF INTERACTION AND L2 APTITUDE ON THE DEVELOPMENT OF THE PRESENT THIRD PERSON SINGULAR THROUGH RECASTS

by

Nektaria-Efstathia Kourtali

PhD Thesis

UCL, Institute of Education, University College London

January 2018

‘I, Nektaria-Efstathia Kourtali confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

ABSTRACT

The role of recasts, a corrective feedback technique, has received much attention from instructed SLA researchers. While a variety of factors have been identified as influencing their effectiveness in promoting L2 development, task complexity, mode of interaction and L2 aptitude are three potential moderator variables that have been the object of relatively little research. To fill these gaps, two studies were conducted. Study 1 investigated the combined effects of task complexity and recasts on modified output and development in knowledge of the present third person singular, the target construction. Study 2 explored the joint impact of mode of interaction and recasts on these outcome variables. Whether L2 aptitude influences these relationships was also examined.

In both studies, 60 young learners, all Greek learners of English, were randomly assigned to one of two experimental conditions. In Study 1, the two groups completed tasks of differential cognitive complexity, with more or fewer reasoning demands. In Study 2, the two groups carried out tasks in different modalities, face-to-face versus computer-mediated. Recasts were provided in response to errors in the target construction. L2 development was gauged by an oral production test and a written production test in both studies, and Study 1 additionally included an elicited imitation test. The LLAMA test (Meara, 2005) was used as an index of L2 aptitude.

In Study 1, students who completed less complex tasks achieved greater gains on the oral and written production tests. In Study 2, the face-to-face group demonstrated greater L2 improvement on both outcome measures. Correlational analyses showed that the gains of the learners who completed more complex tasks were related to learners' ability to recognize oral patterns (LLAMA D) and to associate sounds with symbols (LLAMA E). No correlation was found between aptitude and L2 gains in Study 2.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Corrective Feedback in Oral Interaction

The focus on form approach has inspired a large body of research in the field of instructed second language acquisition. Motivated by Long's Interaction Hypothesis, this approach posits that drawing learners' attention to linguistic elements while engaging in meaningful interaction facilitates subsequent L2 development (Long & Crookes, 1992; Long, 1996, 2000, 2015; Long & Robinson, 1998). One way to promote a focus on form is by providing learners with corrective feedback (CF), i.e. "responses to learner utterances that contain an error" (Ellis, Loewen, & Erlam, 2006, p. 340). A number of recent meta-analyses indicate that CF can assist interlanguage development (e.g. Li, 2010; Lyster & Saito, 2010; Lyster, Saito, & Sato, 2013; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010); however, what types of CF benefit learners more, and under what conditions, remains a disputed issue among SLA researchers. For example, there is disagreement over whether feedback that explicitly draws learners' attention to errors (e.g. explicit correction or metalinguistic information) is more effective than more unobtrusive interventions, such as recasts (example 1.1). There is also a debate about whether feedback that models target-like features (e.g. recasts) differentially affects development in comparison to feedback that encourages learners to correct their errors by themselves (e.g. elicitation and clarification requests) (see Goo & Mackey, 2013 vs Lyster & Ranta, 2013).

Recasts, the focus of the current study, are defined by Long (2007) as:

> A reformulation of all or part of a learner's immediately preceding utterance in which one or more nontargetlike (lexical, grammatical, etc.) items is/are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on *meaning*, not language as object. (p. 77)

The following example (1.1), from the present study, shows how a learner's utterance is reformulated in order to address an error in the present third person singular by employing a recast.

Example 1.1

S. He dance tango

R. He dances? (*recast*)

S. dances tango

From the current study

Previous research has demonstrated that the effectiveness of recasts, in terms of leading to noticing and second language development, is influenced by several variables. For example, an important factor is whether recasts are supplied in a laboratory, in a form-oriented classroom or in a content-oriented, immersion classroom (e.g. Ellis, Basturkmen, & Loewen, 2001;

Han, 2002; Llinares & Lyster, 2014; Lyster & Ranta, 1997; Lyster & Mori, 2006; Lyster & Izquierdo, 2009; McDonough, 2007; Panova & Lyster, 2002; Sheen, 2004). Other moderating factors include learners' level of proficiency (Mackey & Philp, 1998; Philp, 2003; Trofimovich, Ammar & Gatbonton, 2007), learners' educational background (Bigelow, Delmas, Hansen, & Tarone, 2006; Sheen, 2004), learners' age and type of interlocutor (Mackey, Oliver, & Leeman, 2003), type of target linguistic construction (Ellis, 2007; Kartchava & Ammar, 2014; Guchte, Braaksma, Rijlaarsdam, & Bimmel, 2015; Ishida, 2004; Leeman, 2003; Yang & Lyster, 2010; Yilmaz, 2012), type of target error (Kim & Han, 2007; Lyster, 1998; Mackey, Gass, & McDonough, 2000; Smith, 2012; Trofimovich et al., 2007) and the characteristics of recasts (Egi, 2007a; Loewen & Philp, 2006; Nassaji, 2009; Philp, 2003; Roberts, 1995; Sheen, 2006).

1.2 Mode of Interaction, Task Complexity and L2 Aptitude

In addition to the factors presented above, three other variables that may influence the efficacy of recasts are mode of interaction, task complexity and L2 aptitude. Regarding mode of interaction, according to previous meta-analyses, both face-to-face (FTF) and computer-mediated communication (CMC) benefit L2 learners (Li, 2010; Lin, Huang, & Liou, 2013; Ziegler, 2015). Nonetheless, to date, only a limited number of studies have explored synchronous computer-mediated communication (SCMC), and particularly the impact of computer-delivered, written recasts in text-based online chat

(Gurzynski-Weiss & Baralt, 2014a, 2014b; Lai, Fei, & Roots, 2008; Loewen & Erlam, 2006; Sauro, 2009; Smith, 2012; Yilmaz, 2012) (see sections 2.2.2 and 2.2.3). In Loewen and Erlam (2006), neither recasts nor metalinguistic feedback led to more L2 gains in comparison to a control group, while in Sauro (2009), recasts were equally beneficial as metalinguistic feedback. Previous SCMC research has also demonstrated that the effectiveness of recasts in facilitating noticing and subsequent development is influenced by their linguistic focus (Gurzynski-Weiss & Baralt, 2014b; Smith, 2012), the salience of their linguistic target (Yilmaz, 2012) and their contingency to the error (Lai, Fei, & Roots, 2008).

Task complexity refers to the cognitive demands imposed on second language learners when performing a task (Robinson, 2001b; 2003b, 2005, 2011). The current study draws on two cognitive models of task complexity: Robinson's Cognition Hypothesis (2001b; 2003b, 2005, 2011) and Skehan's Limited Capacity Model (2009, 2014b). In Robinson's view, task demands can be increased by two types of task manipulations: (1) those associated with *resource-dispersing* variables (e.g. whether planning time is provided, whether a task is structured etc.) and (2) those related to *resource-directing* variables (e.g. whether a task involves greater reasoning demands, the processing of more elements or present/ past time reference with/ without contextual support).

The predictions of Skehan's and Robinson's models diverge with respect to the effects of increasing the cognitive demands of tasks on L2 outcomes.

According to the Limited Capacity Model, the increased cognitive load of complex tasks (e.g. tasks with more reasoning demands) can make it difficult for learners to achieve both accurate and complex language due to attentional limitations, while the Cognition Hypothesis posits that increasing task complexity is expected to direct learners' attention to linguistic encoding, resulting in more accurate and complex language. Regarding the processing of recasts delivered during complex and simple tasks, different presumptions can be made on the basis of the two models. Specifically, drawing on the Limited Capacity Model, tasks imposing greater cognitive demands may not allow learners to devote sufficient attention to the linguistic area addressed by feedback, as their limited attentional resources are used to meet the enhanced conceptual demands of tasks. In contrast, from the perspective of the Cognition Hypothesis, as learners have to fulfil greater functional and linguistic demands during more complex tasks, they are expected to direct their attention to their interlocutor's input, which may involve CF, such as recasts (see section 2.3.1).

Empirical studies that have examined the efficacy of recasts supplied during tasks manipulated in terms of complexity have demonstrated mixed findings (e.g. Baralt, 2013; Nuevo, 2006; Révész, 2009; Révész, Sachs, & Hama, 2014). For example, Nuevo (2006) found no effects of task complexity, while in Baralt (2014) the mode of interaction (i.e. oral vs computer-delivered recasts) moderated the effects of task complexity on the efficacy of feedback. Moreover, in Révész, Sachs, and Hama (2014), recasts were more beneficial in the simple rather than in the complex condition. Révész (2009) and Baralt (2014) indicate that oral recasts were more effective in promoting learning in what they coded as a complex condition, in comparison to simple tasks. Given these conflicting findings, clearly more research is needed.

L2 aptitude, the third factor investigated in this thesis, has recently witnessed increased attention. Inspired by Carroll's seminal work (1965, 1981), more recent models of L2 aptitude view aptitude as a complex construct, including sub-constructs such as *memory-as-retrieval*, *phonetic coding*, *language analytic ability*, *attentional control* and *working memory* (Skehan, 1998, 2002). Viewing working memory as an aptitude construct has also been an important update since Carroll's work (Linck, Hughes, Campbell, Silbert, Tare, Jackson, & Doughty, 2013; Robinson, 2001a, 2002, 2007a, 2012; Skehan, 1998, 2002). Over the past 15 years, besides revisiting the conceptualisations of aptitude, researchers have begun to explore how aptitude constructs may relate to different cognitive processes within different stages of SLA (Skehan, 2002, 2016; Wen, Biedroń, & Skehan, 2017), and to the effectiveness of different instructional conditions and pedagogical interventions, including focus on form techniques (Robinson, 2001a, 2002, 2007a, 2012).

Drawing on these theoretical developments, studies on aptitude have explored the extent to which it relates to L2 learners' success in different areas of speech production (e.g. pronunciation, fluency, vocabulary and grammar) under instructed conditions (i.e. foreign language classrooms) (e.g., Saito, 2017) and in the ultimate attainment of learners in naturalistic L2 settings (e.g. Abrahamsson & Hyltenstam, 2008; Bylund, Abrahamsson, & Hyltenstam, 2012; DeKeyser, 2000; DeKeyser, Alfi-Shabtay, & Ravid, 2010; Granena, 2014; Harley & Hart, 1997). Other studies have examined the relationship of aptitude with different instructional conditions (e.g. explicit vs implicit) (DeGraaff, 1997; Erlam, 2005; Hwu & Sun, 2012; Robinson, 1997; VanPatten, Collopy, Price, Borst & Qualin, 2013) such as different types of CF (e.g. Li, 2013, 2015; Sheen, 2007; Trofimovich, Ammar, & Gatbonton, 2007; Yilmaz, 2013; Yilmaz & Granena, 2015; Yilmaz, Granena, & Meyer, 2016; Yilmaz & Koylu, 2016). The results of existing research suggest, overall, that L2 aptitude is associated with the quality of L2 learners' speech production, their ultimate attainment and the effectiveness of both implicit and explicit instruction, including implicit and explicit feedback (see section 2.4).

1.3 Focus of the Present Thesis

The present dissertation comprises two studies. The first study examined the impact of recasts and task complexity on the development in knowledge of the present third person singular, while the second study sought to delve into the influence of recasts and mode of interaction on L2 development of the same linguistic construction. In addition, both studies investigated the relationship between L2 outcomes and L2 aptitude. The two research projects also explored whether task complexity/ mode of interaction affected the participants' responses to the feedback they received (i.e. whether they modified their erroneous output by correcting their initial errors after receiving recasts) and whether output modification was linked to L2 development and aptitude.

Although recasts have been the subject of much research attention, the present thesis is novel in several aspects: First, the participants came from an under-researched population, i.e. young learners whose ages ranged from 10.5 to 13 years. Considering that the majority of studies on CF, task complexity, mode of interaction and aptitude have focused on adult learners, the current thesis sought to fill an important gap. Second, in addition to investigating the combined effects of (1) recasts and task complexity and (2) recasts and mode of interaction on L2 gains, the current empirical work explored whether these effects are associated with individual differences in L2 aptitude. Although some cognitive abilities have received researchers' attention, such as WMC (e.g. Goo, 2012, 2016; Kim, Payant, & Pearson, 2015; Li, 2015; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Mackey & Sachs, 2012; Payne & Whitney, 2002; Révész, 2012; Sagarra, 2007a, 2007b; Sagarra & Abbuhl, 2013; Trofimovich et al., 2007; Yilmaz, 2013) and analytic abilities (e.g. Li, 2013, 2015; Sheen, 2007; Trofimovich et al., 2007; Yilmaz, 2013) in CF research, the role of aptitude constructs, such as phonemic coding, in second language learning constitutes an under-explored area. Few studies have looked into its relationship with L2 outcomes (e.g. Yilmaz & Koylu, 2016; Saito, 2017). The two studies of the present thesis also examined learners' development of an under-researched linguistic construction, i.e. the present third person singular. Considering that this target feature is not easily acquired by L2 learners, due to its low salience and communicative redundancy (see section 3.4.3), more research is needed to investigate what learning conditions facilitate L2 development of this feature. Finally, unlike previous studies that have mainly focused on more implicit recasts (i.e. full recasts that reformulate learners' whole utterances), the current thesis employed more explicit recasts that were partial and pinpointed the erroneous parts of learners' output (see section 2.1.3 regarding explicit vs implicit recasts).

1.4 Thesis Outline

The remainder of the present dissertation is organized as follows: A literature review is conducted in the second chapter. Chapter 3 reports on the first study. It begins with a pilot study followed by an overview of the design, and it includes information about the participants, the linguistic target, the type of feedback delivered and the research materials employed (e.g. treatment tasks, outcome measures and questionnaires). In addition, a detailed explanation of the methodology (e.g. coding procedure and statistical analysis) is provided. After the methodology, the results are presented, followed by a discussion of the findings for each research question. Chapter 4 focuses on the second study and has a similar structure to Chapter 3 (study 1). Finally, Chapter 5 draws some conclusions, including a discussion of the theoretical and practical implications of the two studies and suggestions for future research.

CHAPTER 2: REVIEW OF THE LITERATURE

This chapter is divided into four sections. The first section discusses pedagogical approaches involving synthetic and purely analytic syllabi prior to a description of the focus-on-form approach. Then, it presents the theoretical rationale for the focus-on-form approach (e.g. the Interaction Hypothesis and the Noticing Hypothesis), followed by a discussion of how the approach may be put into practice with a focus on the CF literature. First, different types of CF and their characteristics are presented, before the findings of research on CF are discussed in detail. The focus is on (1) studies that explored noticing of CF and (2) those that gauge L2 development resulting from CF by employing a pretest-posttest design.

The second, third and fourth sections of the literature review are devoted to previous literature investigating the three main variables of the current dissertation: mode of interaction, task complexity and L2 aptitude. In the second section, several potential advantages of the SCMC mode are presented, followed by studies examining the effects of CF delivered in an SCMC and/or FTF environment. The third section begins with a description of the theoretical framework of task-based language teaching (TBLT) and explains two cognitive models of TBLT (i.e. the Cognition Hypothesis and the

Limited Capacity Model). A discussion of these models is followed by a review of empirical research related to the effects of task complexity on the efficacy of CF. Finally, the fourth section is devoted to a description and discussion of previous research on individual differences. A variable referring to individual differences that has received much attention in previous research is working memory capacity (WMC). After a brief presentation of studies on WMC and recasts, the fourth section mainly focuses on L2 aptitude. It begins with a theoretical review of aptitude research from Carrol to more recent years, and it is followed by empirical studies investigating the relationship of L2 aptitude with learners' achievements/ ultimate attainment and the extent to which L2 aptitude is linked to the effectiveness of different instructional conditions and types of CF techniques.

2.1 Focus-on-Form Approach and Corrective Feedback

2.1.1 Analytic versus Synthetic Pedagogical Approaches

Traditionally, a distinction is made between analytic and synthetic pedagogical approaches. According to synthetic approaches, explicit instruction of linguistic elements plays a pivotal role in second language learning. The focus of *synthetic approaches* is the *language* that should be taught and the transmission of declarative knowledge of linguistic units. The order and timing of presentation of these units is based on a pre-set syllabus (i.e. a lexical, grammatical or notional-functional syllabus). Synthetic approaches also involve the Presentation-Production-Practice (PPP) of artificial classroom language, which often involves inauthentic and simplified language that is presented in texts or produced by learners during oral exercises (for a review, see Long, 2015).

Notwithstanding the widespread use of synthetic syllabuses, the main criticism they have received is associated with their assumption that L2 learners should learn what they are taught in an order determined by a synthetic syllabus. Put differently, the approach assumes that learners start from zero L2 knowledge and then, as they are taught linguistic constructions in isolation, they learn them in a linear way. But SLA researchers have pinpointed two main problems with this view. First, the majority of linguistic constructions are not learnt in isolation. For example, Long (2015) explains that a synthetic syllabus might present how tenses change from active to passive voice outside a real-life context; however, in order to produce an utterance like “The gift wasn’t bought by John in Paris”, learners should control various aspects of the language, such as word order, tense, auxiliaries in the passive, negation and prepositions.

A second problem with synthetic approaches is that, according to previous literature, L2 learners do not learn in a linear way, as synthetic syllabuses assume. That is, learning of new linguistic constructions does not occur in harmony with a pre-determined syllabus but follows an internal *learner syllabus* (Corder, 1967). Previous research has demonstrated that new linguistic features are learnt only when learners are developmentally ready, despite teachers' efforts to control and change the timing and order in which different linguistic areas are acquired (Pienemann, 1984, 1989; Pienemann & Kessler, 2011). For example, learners only produce negation and question formation accurately after they have passed through specific developmental stages (see Pienemann, 1989). It should be noted that the notion of developmental sequences has been a matter of recent debate among SLA researchers (e.g. see Lenzing, 2015; Lowie & Verspoor, 2015; Pienemann, 2015; Zhang & Lantolf, 2015).

To resolve the problems posed by synthetic syllabi, more contemporary pedagogical approaches, referred to as *analytic,* have been suggested by SLA researchers. Unlike synthetic approaches, *analytic approaches* focus on meaning rather than form and they entail meaningful L2 communication in a more natural and authentic context. They also put more emphasis on the *learner* rather than on the *language*. Learners are expected to utilize innate cognitive abilities to analyse the input and infer grammatical rules while they engage in tasks or activities that encourage holistic use of the second language (Long, 1985; Long & Crookes, 1992, Long & Robinson, 1998; Wilkins, 1976). Two well-known approaches that adopt analytic syllabi are the communicative approach and task-based language teaching (see section 2.3).

Although there is consensus among SLA researchers that analytic approaches appear to be more effective than synthetic approaches in terms of leading to L2 outcomes, many researchers advocate that native-like proficiency cannot be achieved when a purely analytic approach is employed (e.g. Skehan & Foster, 2001; Long, 2015). They contend that analytic approaches are more advantageous for L2 learners when they are supplemented by explicit instruction of formal properties of the second language. The extent to which a more explicit focus on language is needed is related to several factors, such as: (1) learners' age, (2) pace of learning and amount of L2 exposure in educational contexts, (3) tendency of errors to fossilise and (4) salience of linguistic constructions. Now we turn to a brief discussion of each of these factors.

First, regarding learners' age, some researchers posit that older learners and adults benefit to a greater extent from an analytic approach when it integrates explicit instruction that draws their attention to form (e.g. DeKeyser, 2003; Paradis, 2004, 2009). Second, when a purely analytic syllabus is implemented, the process of inferring rules might require time; and in an EFL context whereby exposure to the second language is limited, more explicit interventions might be necessary in order to speed up the learning process (DeKeyser, 2003; Long, 2015). Long (2015) explains that although such explicit interventions might not change learners' internal syllabi and the stages of developmental readiness, as a synthetic syllabus predicts, they might still be useful as they have the potential to make interlanguage development faster and to facilitate ultimate L2 attainment (Long, 1983, 2015). Third, a purely analytic approach that is devoid of attention to form might result in the fossilization of errors, i.e. permanent non-target-like production of linguistic constructions (Skehan & Foster, 2001). For example, purely content-based instruction in French immersion programmes in Canada was only partially beneficial. Although the learners achieved L2 comprehension at native levels, the lack of explicit focus on formal properties of the L2 during the programme negatively influenced the learners' grammatical competence during L2 production (Swain, 1991).

Finally, it should be emphasized that a solely analytic content-based approach might impede the acquisition of non-salient linguistic constructions. Perceptual salience refers to features that constitute a linguistic construction in a way that is "more visually or auditorily prominent than another" (Dulay, Burt, & Krashen, 1982, p. 33). The salience of linguistic features has been attributed to several factors, such as phonetic substance, stress, pitch and position (Leeman, 2003; Slobin, 1973). For example, in oral interaction, salient constructions are easily heard and noticed when they are sonorous, stressed and produced with high pitch (Leeman, 2003). With regard to position, salience increases when morphemes are free rather than bound, or when a construction is found at the end of a word or utterance (Bardovi-Harlig, 1987; Leeman, 2003). For example, Bardovi-Harlig (1987) found that preposition stranding (e.g. "Who did Bob report the accident *to*?", "The woman who Bob sent a postcard *to* was his aunt.") was acquired *before* preposition pied-piping (e.g. "*For* whom did Diane bake a cake?", "The teacher helped the student *for* whom the lesson was difficult.") in both dative wh-questions and relative clauses. Bardovi-Harlig explains that preposition stranding is more salient than preposition pied-piping due to its position in the utterance.

According to previous literature, other features associated with salience are frequency, grammatical/ syntactic and semantic complexity, communicative value and the similarity between L1 and L2 (Bardovi-Harlig, 1987; Goldschneider & DeKeyser, 2001; Larsen-Freeman & Long, 1991; VanPatten, 1996; Yilmaz, 2012). Empirical studies suggest that infrequent, complex and communicatively redundant constructions are more difficult to acquire than their counterparts. For example, Bardovi-Harlig (1987) used two linguistic constructions that differed in syntactic complexity: relative clauses and one-clause dative questions, with the former being more syntactically complex than the latter. Interestingly, Bardovi-Harlig found that non-salient preposition pied-piping was acquired faster in a syntactically simple construction in comparison to its counterpart in a syntactically complex construction.

Overall, when non-salient constructions with the features described above (e.g. perceptual salience, lack of frequency and communicative value, and grammatical/ syntactic/ semantic complexity) are not explicitly addressed by content-based analytic syllabi, they might remain unattended to by L2 learners. Hence, an analytic approach that mainly focuses on meaning rather than form might hamper the development of non-salient linguistic items and so a more explicit focus may be necessary to draw learners' attention to such features (for a review on salience, see Goldschneider & DeKeyser, 2001).

To resolve the above problems of purely analytic syllabi related to learners' age, educational context in tandem with length of exposure to the L2 and the salience of linguistic elements, learners' attention should be diverted from meaning to form during L2 interaction in either a more explicit or more implicit/ unobtrusive manner. To this end, Long suggested a novel pedagogical approach, namely, an *analytic approach with a focus on form* (Long & Crookes, 1992; Long, 1996, 2000, 2015; Long & Robinson, 1998). Due to the immense influence of this approach on instructed SLA research, including the current study, it is discussed in greater detail in the following section.

2.1.2 Analytic Approaches with a Focus on Form

Long (2015) argues that both a synthetic approach that entails a *focus on forms* and an analytic approach that involves only a *focus on meaning* are accompanied by several flaws, as explained in the previous section. To eradicate their drawbacks, Long advocates an alternative pedagogical option, i.e. an analytic approach with a *focus on form* (Long & Crookes, 1992; Long, 1996, 2000, 2015; Long & Robinson, 1998). The focus-on-form approach recognizes that the cognitive processes associated with noticing and attention to linguistic constructions in the L2 input play a crucial role in second language learning. The importance of these processes is also emphasized by the Noticing Hypothesis formulated by Schmidt (1990, 1995, 2001). Before discussing the focus-on-form approach and its practical implications in the ELT classroom, the theoretical underpinning that has to some extent motivated it, i.e. the Noticing Hypothesis, will be discussed in the current section.

Schmidt (1990, 1995, 2001) claims that a conscious cognitive process, namely, noticing involves learners' attention to the input and is a prerequisite for second language learning. Schmidt contends that the process of noticing implicates two psychological mechanisms: focal attention and a low level of awareness. The important role of attention to L2 input in second language learning has also been emphasized by several SLA researchers (e.g. Gass, 1990; N. Ellis, 2004). However, there is disagreement as to whether awareness is required during attentional processes in order to facilitate L2 development. Schmidt (2001) argues that "a low level of awareness, called here "noticing," is nearly isomorphic with attention and seems to be associated with all learning" (p. 1). He also underlines that noticing is related to "awareness only at a very low level of abstraction" (Schmidt, 2001, p. 5) and that a higher level of awareness than noticing, that entails understanding of rules, might assist L2 learning but is not necessary. Opposite to Schmidt's argument, Tomlin and Villa (1994) state that L2 *detection* of linguistic elements during attentional processes is sufficient for second language learning, and detection does not involve awareness. Schmidt (2001) and Robinson (2003a) agree with Tomlin and Villa about the crucial role of detection in second language learning;

however, they point out that detection devoid of awareness is not sufficient for L2 development. In other words, the part of input that is detected by learners has to be consciously registered in order to become *intake* (i.e. to be further processed so as to be learnt).

Several researchers also emphasize the pivotal role of memory processes while noticing and attending to L2 input (N. Ellis, 2005; Long, 2015; Robinson, 2003a). Long (2015) explains that as a result of noticing and attention, form-meaning mappings are processed and rehearsed, while temporarily remaining in short-term memory, until an initial representation of these mappings is formed and stored in learners' long-term memory (see section 2.4). Long (2015) adds that this can facilitate input processing that takes place at the lower level of *detection*. During detection, learners detect form-meaning mappings in the input, but without employing focal or selective attention and without being aware of it. However, Long argues that learners do detect and register form-meaning mappings, when these mappings have already been noticed by learners before and have already been stored in long-term memory. Thus, noticing and attending to a linguistic feature at a higher level creates an initial representation of it in long-term memory, and this representation is reinforced when, at a lower level, learners detect a similar feature in the input. In other words, the initial representation of a linguistic item primes L2 learners to detect and perceive similar items in the input.

The focus-on-form approach also considers attention to linguistic features in the input to be necessary for L2 learning (Long & Crookes, 1992; Long, 1996, 2000, 2015; Long & Robinson, 1998). Long (2000) explains that:

> Focus on form refers to how attentional resources are allocated and involves briefly drawing students' attention to linguistic elements … *in context,* as they arise incidentally in lessons whose overriding focus is on meaning, or communication. The temporary shifts in focal attention are triggered by students' problems with comprehension or production. (p. 185)

Doughty and Williams (1998) propose a slightly different implementation of the focus-on-form approach. Although they agree with Long that learners' attention should be devoted to form while engaging in meaningful interaction, their proposal diverges from Long's perspective with regard to the incidental nature of focus on form. As explained earlier, Long (2000) argues that learners should attend to linguistic constructions "as they arise incidentally in lessons" (p. 185). In contrast, Doughty and Williams additionally suggest the identification of specific linguistic problems of learners and planned interventions in order to address them. Doughty and Williams's proposal can be of great pedagogical value for non-salient linguistic constructions that usually remain unattended to during interaction. As the current study only targets a single, non-salient linguistic feature, i.e. the present third person singular (see section 3.4.3), the focus-on-form approach is applied following Doughty and Williams's formulation.

2.1.3 Corrective Feedback

The focus-on-form approach subsumes pedagogical interventions, such as input enhancement or the provision of corrective feedback (CF). The benefits of CF are largely undisputed by SLA researchers, as previous meta-analyses have demonstrated its effectiveness in promoting L2 development (e.g. Li, 2010; Lyster & Saito, 2010; Lyster, Saito, & Sato, 2013; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010). Nonetheless, one of the thorniest issues in the field of instructed SLA is clarifying what types of CF, and under what learning conditions, are more beneficial for L2 learners. A plethora of research has been conducted in order to shed light on these questions. Before discussing the results of empirical studies regarding the effectiveness of CF, the following section presents different types of CF and their characteristics.

Types of Corrective Feedback

One of the most common focus-on-form techniques used as CF is recasts (see section 1.1 for a definition of recasts and an example). Others that have also received attention from SLA researchers are prompts, explicit correction and metalinguistic feedback. Prompts involve *elicitations*, *repetitions* and *clarification requests*. Based on Lyster and Ranta's (1997) categorization, prompts can take the following forms: *Elicitation* "pushes" learners to correct their erroneous utterances on their own when (1) the interlocutor (e.g. teacher) repeats part of their utterances and pauses, just before erroneous parts, so that learners can continue with self-repairs (i.e. self-correction after teacher feedback), (2) the teacher attempts to elicit target-like construction by asking questions such as "How do we say X?", (3) the teacher requires learners to reformulate their utterances. Similarly, *clarification requests* are also attempts to elicit self-repairs from learners by asking them questions such as "What do you mean by X?". As for *repetitions*, the teacher repeats learners' errors, aiming at self-repair. These techniques can be employed in combination with other types of feedback (e.g. metalinguistic feedback followed by elicitation). *Metalinguistic feedback* involves information about errors using metalanguage (e.g. grammatical terms such as "past tense") or comments/ questions that explicitly require learners to identify errors (Lyster & Ranta, 1997). *Explicit correction* entails the teacher clearly informing students about their non-target-like output and perhaps also providing them with correct alternatives (e.g. X is not correct).

Explicit vs Implicit Corrective Feedback

According to Li (2010):

> Implicit feedback refers to any corrective move that does not overtly inform the learner of the unacceptability of his/her erroneous production; explicit feedback, in contrast, draws the learner's attention to the error he/she commits. (p.337)

In general, explicit correction and metalinguistic feedback are considered explicit techniques, whereas prompts and recasts are usually categorised as implicit moves (Li, 2010). However, such a division can be problematic for recasts. Although recasts are portrayed as an implicit focus-on-form technique due to their unobtrusiveness (Long & Robinson, 1998), Mackey and Goo (2007) rightly argue that recasts are "elastic in nature" (p. 413). A strict dichotomy that classifies recasts as implicit intervention ignores certain characteristics that constitute them as a polymorphous type of feedback. Sheen (2006) points out that recasts lie on an implicit-explicit continuum, and whether they are towards the explicit or implicit end is determined by several characteristics. Previous research has attempted to identify these characteristics in order to explore their impact on L2 learning (Braidi, 2002; Ellis & Sheen, 2006; Loewen & Philp, 2006; Lyster, 1998; Philp, 2003; Sheen, 2006).

First and foremost, Sheen (2006) differentiates multi-move recasts from single-move recasts. The former involve (1) repetition of an error followed by a recast (see Doughty & Varela, 1998), (2) repeated recasts that reformulate learners' errors either fully or partially, and (3) recasts combined with other types of feedback such as metalinguistic information. A single-move recast refers to a single turn that includes a recast (i.e. full or partial reformulation of a learner's error). Single-move recasts may differ in their characteristics. Lyster (1998) distinguishes recasts based on *mode* and *scope*. *Mode* refers to whether a recast is *declarative* or *interrogative*, and *scope* is associated with whether a recast is *isolated* or *incorporated*. Unlike isolated recasts, incorporated ones are part of longer statements that provide or ask for additional information.

Additional features which can distinguish recasts include *prosodic emphasis, reduction, length, number of changes, type of change* and *linguistic focus* (e.g. Philp, 2003; Sheen, 2006; Loewen & Philp, 2006). *Prosodic emphasis* refers to whether a word or morpheme is stressed (Chaudron, 1977; Sheen, 2004, 2006; Loewen & Philp, 2006). As for *reduction*, Sheen (2006) explains that it is associated with whether the reformulation of a recast is shorter than a learner's non-target-like output as opposed to recasts that reformulate a learner's whole utterance. In previous literature, recasts with reduction have also been called *partial recasts* (Roberts, 1995), or recasts with *segmentation* as they segment the erroneous parts of learners' utterances (Loewen & Philp, 2006).

With regard to their *length*, recasts have been coded as those that (1) include only one word or short phrase and those whose reformulation includes longer phrases or clauses (short vs long recasts) (Sheen, 2006). For example, in Philp's (2003) study, recasts with fewer than five morphemes were coded as short. Another important feature is the *number of changes* (Braidi, 2002; Sheen, 2006). This is related to the number of linguistic features that recasts seek to correct. Recasts are categorized into those that aim at one change (i.e. they address only one non-target-like linguistic item) and those that involve multiple changes (i.e. their goal is to correct more than one error).

Recasts are also differentiated with respect to the *type of change* they make (Braidi, 2002; Sheen, 2006). Type of change includes *addition, deletion, substitution* and *reordering. Addition* occurs when a recast adds a linguistic feature missing from the learner's output; d*eletion* takes place when a recast removes a feature; *substitution* entails the replacement of a non-target-like feature with a correct one; and finally, *reordering* involves modification of the order of linguistic features. Finally, recasts vary in their *linguistic focus* depending on the types of errors they target (e.g. pronunciation, semantic, lexical or morphosyntactic) (Lyster, 1998; Mackey, Gass, & McDonough, 2000; Sheen, 2006; Smith, 2012).

The features described above might transform recasts into more explicit or implicit feedback. For example, partial, reduced recasts combined with prosodic emphasis (e.g. stressing a learner's error) are more explicit than longer recasts provided as part of a conversation. Distinguishing explicit from implicit recasts is crucial, as explicit recasts might have a different influence on second language learning than implicit ones. For example, explicit recasts might be more effective in drawing learners' attention to non-salient linguistic constructions. However, it should be noted that salience and explicitness are two different constructs. Sheen (2006) explains that salience is a psycholinguistic construct associated with how easily a linguistic item is noticed, whereas the explicitness of feedback is related to discourse characteristics. For example, more explicit feedback that involves metalinguistic information or prosodic stress is a discourse phenomenon that may increase the salience of the linguistic target.

In summary, different types of CF differ as to whether they constitute more explicit or implicit moves during interaction. In particular, previous SLA researchers classify explicit correction and metalinguistic feedback as explicit interventions and prompts as more implicit focus-on-form techniques. With respect to recasts, they may be categorized as more explicit or implicit feedback, depending on their characteristics. Apart from their explicitness vs implicitness, another important aspect that differentiates types of CF is whether they involve positive or negative evidence. This distinction is discussed in the following section.

Positive vs Negative Evidence

Gass (1997, 2003) explains that positive evidence involves the provision of target-like exemplars in the input that indicate whether a construction is possible in the L2. In contrast, the focus of negative evidence is what cannot be used in the L2, by correcting or explaining non-target-like constructions. In other words, the aim of negative evidence is to draw learners' attention to their errors, whereas positive evidence puts more emphasis on a target-like model. CF moves have also been distinguished by Ellis (2009) into *output-prompting* and *input-providing*. These two types of feedback differ as to whether they focus on the provision of negative evidence or whether they involve positive evidence that models target-like linguistic constructions. An example of output-*prompting* CF is prompts, as they "push" learners to correct their non-target-like utterances on their own. Unlike recasts, prompts (i.e. elicitations, metalinguistic clues, clarification requests or repetitions) offer only negative evidence rather than positive evidence. That is, they do not model a target-like construction but intend to draw learners' attention to their errors. Lyster (1998) contends that output-pushing CF, such as prompts, is highly beneficial for L2 learners as it encourages them to retrieve already existing target-like constructions from long-term memory on their own. According to Lyster, as learners employ their own cognitive resources, they engage in deeper processing of the linguistic target. Nonetheless, prompts cannot lead to the encoding of new knowledge and self-repairs cannot be achieved for linguistic elements that have not already been internalized by learners (Long, 2007). Hence, input-providing CF might be more effective as regards leading to knowledge of novel constructions.

Recasts are classified as *input-providing* CF (Ellis, 2009), as they involve both negative and positive evidence. That is, recasts aim to signal to learners that they have produced a non-target-like utterance and, simultaneously, model a target-like construction. Several advantages of recasts have been attributed to this characteristic. First, the provision of a target-like construction immediately after a learner's non-target-like utterance might encourage them to engage in making a cognitive comparison between their erroneous output and the target-like model provided by the recast (Doughty, 2001; Long, 2007). Second, as recasts model the target-like construction, they might generate new knowledge and lead to the creation of initial mental representations of linguistic features. As explained earlier, this cannot be achieved by other types of CF that rely on self-repairs. They are restricted in prompting learners to employ their internal resources in order to retrieve linguistic features that have already been encountered before and which have been internalized to some degree (Long, 2007). Third, recasts involve reformulations of learners' own utterances. As learners already know the meaning they intend to convey, they may free up attentional resources and allocate them to the target linguistic model provided by the recast. In particular, Long (2007) explains that:

> …recasts convey needed information about the target language *in context*, when interlocutors share a *joint attentional focus*, and when the learner already has *prior comprehension* of at least part of the message, thereby facilitating form-function mapping. (p. 77)

Long adds that:

> …learners are *vested* in the exchange, as it is their message that is at stake, and so will probably be *motivated* and *attending*, conditions likely to facilitate *noticing* of any new linguistic information in the input. (pp. 77–78)

Finally, an important advantage of recasts is that they do not treat formal properties of the second language in isolation but rather as part of a meaningful interaction. Hence, as recasts divert learners' attention to a model form during interaction that mainly focuses on meaning, they promote the creation of form-meaning mappings (Doughty, 2001). It is also worth noting that all of the advantages presented above are achieved in an unobtrusive manner, and which is less face-threatening for learners in comparison to other types of feedback such as explicit correction (Doughty, 2001).

In line with the above arguments, two empirical studies have demonstrated that the effectiveness of recasts in promoting L2 development is associated with their function as positive evidence (i.e. they provide a target-like linguistic model), rather than their function as negative evidence (i.e. signalling that a non-target-like feature was produced by the learner) (Egi, 2007b; Leeman, 2003). Leeman (2003) found that the most effective conditions for promoting L2 development were (1) implicit recasts that involved a combination of negative evidence and positive evidence, and (2) only positive evidence with enhanced salience by using intonation and additional stress. On the contrary, conditions that did not facilitate learning were (1) negative evidence alone (i.e. implicitly informing the learners that they produced a non-target-like feature without providing a target-like model) or (2) positive evidence without enhanced salience. Similarly, Egi (2007b) found that learners benefited from recasts only when they interpreted them as positive evidence (i.e. they noticed a target-like construction but they did not report noticing their error) or when they interpreted recasts as both positive and negative evidence (i.e. they noticed their error and the target-like model provided by a recast). In contrast, they did not demonstrate any significant improvement when they perceived recasts only as negative evidence (i.e. learners recognized their errors but they did not report noticing the target-like model provided by a recast).

Leeman's and Egi's findings indicate that in order for learners to exhibit L2 gains after receiving recasts, it is not sufficient to draw their attention to their errors by offering only negative evidence. CF should also encourage them to attend to the target-like linguistic model. Interestingly, in Leeman's study, the positive evidence was more beneficial under the condition where a target model was provided with enhanced salience. This indicates that stress and intonation also appear to have a facilitative role in L2 learning.

Previous empirical studies have also compared the effectiveness of different CF techniques (e.g. explicit feedback vs recasts or prompts vs recasts) leading to L2 benefits. These studies can be divided into (1) those that explored the potential of CF moves to facilitate the noticing of target linguistic constructions, and (2) those that examined the extent to which different types of CF promote subsequent L2 development. The following section discusses studies that focused on noticing.

2.1.4 Studies on Corrective Feedback and Noticing

As explained earlier, according to Schmidt's (1990, 1995, 2001) Noticing Hypothesis, noticing linguistic features in the input is a necessary condition for L2 development. Drawing on Schmidt's argument, a group of researchers set out to explore whether different types of CF were effective in assisting noticing of their linguistic targets.

An initial attempt to measure noticing involved examining learners' responses to the feedback they received. Based on Lyster and Ranta's (1997) coding, these responses, namely *uptake,* are distinguished into uptake with *repair* and uptake that *needs repair*. The former is viewed as successful as its production entails learners correcting their initial errors and, according to Lyster and Ranta (1997), may involve *repetition*, *incorporation*, *self-repair* or *peer repair*. *Input-providing* feedback, such as recasts, generates *repetitions* of a target-like feature or its *incorporation* into a longer utterance. *Self-*

repairs result from *output-pushing* feedback, such as prompts, and they occur when students are encouraged to correct target errors on their own. *Peer repair* takes place when the target error is corrected by a student other than the one who produced the non-target-like item. As uptake with repair entails learners modifying their initial non-target-like output, researchers have also named it *modified output* (e.g. Gurzynski-Weiss & Baralt, 2014a; Mackey & Philp, 1998; Mackey, Oliver, & Leeman, 2003; Oliver & Mackey, 2003). In the current study, when learners correct an initial error after receiving a recast, the more general term uptake with modified output will be used (see section 3.4.11). Unlike uptake *with repair* or *modified output*, uptake that *needs repair* is produced when students respond to feedback without correcting their erroneous utterances. According to Lyster and Ranta (1997), learners' responses need repair when they simply reply yes or no, when they make the same error as the first one or a new one, when they correct only part of the initial error or when they hesitate.

The extent to which *input-providing* CF in the form of recasts is effective in producing modified output has received great attention from SLA researchers; however, previous empirical studies have demonstrated contradictory findings. These conflicting results have been attributed to several factors related to educational or interactional context and the characteristics of recasts. Educational context refers to the type of instruction (e.g. content-based vs form-based classrooms) and learner characteristics. Interactional context refers to the main focus of an activity or task when feedback is provided (e.g. meaningful communication or a focus on certain linguistic features). The characteristics of recasts are related to certain features that transform them into more explicit or more implicit feedback (see section 2.1.3). These factors are discussed below.

Factors Affecting the Noticing of Recasts

Contextual factors

With regard to type of educational setting, it has been suggested that recasts might fail in promoting the noticing of linguistic features during content-based lessons whereby learners engage in highly meaning-oriented activities. For example, in a study conducted in French immersion classrooms in Canada, Lyster and Ranta (1997) observed that although recasts were the type of feedback most frequently employed, in comparison to elicitations, clarification requests, repetitions, metalinguistic feedback and explicit correction, they led to lower amounts of modified output than other types of feedback, especially elicitations. In a similar vein, Panova and Lyster (2002) showed that recasts were the most common CF technique; however, they generated significantly lower rates of modified output in comparison to elicitations and repetitions. Nevertheless, unlike Lyster and Ranta's (1997) and Panova and Lyster's (2002) findings, other studies have exhibited contradictory results. For example, Ellis, Basturkmen, and Loewen (2001), Sheen (2004) and Llinares and Lyster (2014) found that the majority of recasts were followed by modified output in an ESL class in New Zealand, in a Korean EFL class and in a CLIL classroom in Spain, respectively.

Several researchers have attributed these conflicting findings to differences in the type of educational setting (Llinares & Lyster, 2014; Lyster & Mori, 2006, Sheen, 2004). In particular, Lyster and Mori (2006) compared the amount of modified output generated by recasts in two types of educational contexts: French immersion classrooms in Lyster and Ranta's (1997) study and Japanese immersion classrooms. They found that in the latter, recasts were more successful in promoting the production of modified output in comparison to French immersion classrooms. Lyster and Mori accounted for this difference in terms of learners' orientation; the French immersion classrooms were meaning-oriented whereas the Japanese ones were form-oriented. Lyster (1998) explains that, in meaning-oriented classrooms, a focus on content outweighs a focus on form, and consequently, recasts might be misinterpreted by learners as conversational moves that function as confirmation of the semantic content of their utterances. As a result, recasts fail to divert learners' attention from meaning to form and the target linguistic feature remains unattended to. Furthermore, during content-based lessons, many recasts are followed by topic continuation without providing learners with an opportunity to correct their initial errors (e.g. in Lyster & Ranta's study).

In contrast, the intention of recasts, i.e. to draw learners' attention to form, is less ambiguous in other educational settings, such as the Korean EFL classroom (Sheen, 2004), the ESL classroom in New Zealand (Ellis et al., 2001) and the Japanese classroom (Lyster & Mori, 2006). In such contexts, students appear to be more "trained" to prioritize attention to form-meaning mappings rather than meaning alone. Lyster and Mori (2006) explain that in such form-oriented educational contexts, the learners perform more form-based activities with a focus on accuracy (e.g. choral repetitions of isolated linguistic constructions). In other words, the students are primed to pay attention to form, thus they are more likely to perceive recasts as focus-on-form episodes and to repeat the target-like construction. Interestingly, even in CLIL classrooms, which integrate content and language, recasts can promote modified output when they are used as didactic rather than conversational moves (Llinares & Lyster, 2014).

Therefore, recasts are more successful in facilitating the noticing of a linguistic element when their intention to promote a focus on form during meaningful interaction is clear to learners (Nicholas, Lightbown, & Spada, 2001). In contrast, if recasts are perceived as semantic comments, they are less likely to lead to noticing and L2 development of the target linguistic construction. Using data based on learners' perceptions of CF via immediate reports and stimulated recalls (see Gass & Mackey, 2000), Egi (2007b) found that learners demonstrated significantly lower subsequent L2 gains when they reported that they attended only to the semantic content of recasts or when they had perceived them as confirmation of meaning, and thus they had ignored the target linguistic problem.

Apart from the educational context, the interactional context may also influence the efficacy of recasts. In particular, learners might be more prone to erroneously perceiving recasts as conversational moves rather than focus-on-form techniques when the feedback occurs during interactional contexts that prioritize a focus on content rather than on language. For example, Oliver and Mackey (2003) identified the provision of CF in four interactional contexts that differed as to whether their primary focus was: (1) content (e.g. asking students to name animals in pictures), (2) management (e.g. giving instructions for tasks), (3) communication (e.g. conversation) and (4) explicit language (e.g. activities focusing on formal aspects of language). The researchers found that during an interactional context that involved explicit language-focused exchanges, recasts and explicit correction were the dominant type of CF and learners also produced the greatest amount of modified output in comparison to the other three interactional contexts. The researchers concluded that in interactional contexts that involve more form-focused activities, learners are more likely to attend to the linguistic area targeted by recasts and their attention is successfully shifted from meaning to form-meaning mappings, as they expect to be corrected.

Other contextual factors that might be related to different rates of modified output across educational settings refer to the characteristics of the recipients of CF (e.g. learners' educational background, ages and amount of acquired explicit knowledge). With respect to educational background, Sheen (2004) noted that classrooms where recasts induced more modified output (i.e. EFL and ESL classrooms in Korea and New Zealand) contained educated adults, whereas recasts were less successful in settings where the students were children or less educated adults (i.e. French immersion classrooms). Sheen (2004) presumed that learners' different responses to CF might be related to their previous learning experiences. In particular, adults with a stronger educational background might be more likely to attend to the target linguistic model of recasts and modify their output in comparison to children and less educated adults. Sheen's argument is in line with Bigelow, Delmas, Hansen, and Tarone's (2006) findings. The participants in their study were divided into two groups that differed in terms of their L1 and L2 literacy levels. The researchers demonstrated that the more literate students completely or partially corrected their initial errors after receiving recasts significantly more than the less literate ones. The difference between the two groups was greater for recasts that targeted two or more errors.

Learners' ages in combination with the type of interlocutor (i.e. native speaker vs non-native speaker) is another factor that might influence the rate of modified output produced by CF including recasts. Mackey, Oliver, and Leeman (2003) found that feedback was followed by greater amounts of modified output when it addressed children whose ages ranged from 8 to 12 years old rather than adults, especially when young learners interacted with non-native speakers. The researchers also underscored the importance of providing learners with opportunities to modify their output. Finally, Sato (2011) has argued that learners with greater explicit knowledge might be more likely to notice recasts as they are already aware of the target linguistic construction prior to the provision of implicit feedback. Hence, it might be easier for students to recognise the linguistic target of recasts.

Characteristics of recasts

Apart from the contextual factors presented above, the characteristics of recasts might also impact on the amount of modified output. For example, Sheen (2006) found that the characteristics of single-move recasts, such as length, reduction, type of change, linguistic focus, mode and number of changes, are associated with the extent to which learners produce modified output. In particular, in Sheen's study, short recasts (i.e. those involving one word or a short phrase) and reduced recasts (i.e. shorter than learners' erroneous utterances) generated more modified output than longer recasts (i.e. clause-length recasts). Sheen's finding aligns well with research that measured noticing by employing retrospective data (e.g. immediate or stimulated recalls) and asking participants to report how they interpreted recasts during interaction (Egi, 2007a; Phil, 2003; Roberts, 1995). In all of these studies, short recasts were accurately recalled by learners to a greater extent than long recasts. For example, in Egi (2007a), learners reported that they interpreted short recasts as negative and positive evidence. In contrast, they perceived longer recasts as responses related to the content of their utterances. Interestingly, this finding was the same for both lexical and morphosyntactic recasts. Regarding the type of changes made by recasts, Sheen (2006) found that *substitution* led to more modified output than *addition* and *deletion*. Sheen points out that replacing one or more linguistic items might be more salient, and consequently more noticeable to learners, than adding or removing linguistic features.

With respect to linguistic focus, in Sheen's study, when the target error was pronunciation, recasts produced more modified output than when the error was grammatical or lexical. Similarly, Lyster (1998) found that phonological errors were modified to a greater degree than grammatical ones after the provision of recasts. Other studies have also demonstrated that learners reported less noticing for feedback that addressed morphosyntax in comparison to other types of errors (e.g. lexical or phonological) (Kim & Han, 2007; Mackey et al., 2000; Smith, 2012; Trofimovich et. al., 2007). With respect to the mode of recasts, Sheen (2006) showed that interrogative recasts were followed by greater amounts of uptake (with and without modified output); however, declarative recasts were more effective in producing output modification. Sheen noted that when recasts are delivered as questions, learners are encouraged to respond and produce uptake, but not necessarily to notice and correct their errors.

The number of changes made by recasts also seems to influence the amount of modified output. In particular, one-change recasts (i.e. recasts that target only a single erroneous linguistic element) induced more modified output than those making several changes in Sheen (2006). In studies that used immediate or retrospective comments of learners regarding their perceptions of feedback, recasts that made more than one or two changes were less accurately perceived than those that addressed only one or two changes (Egi, 2007a; Kim & Han, 2007; Philp, 2003). For example, in Egi (2007a), learners were more likely to interpret recasts with more than two changes as responses to the semantic content of their utterances without attending to their linguistic focus, regardless of whether the recast was morphosyntactic or lexical.

Overall, previous research indicates that contextual factors and the characteristics of recasting are associated with the extent to which learners modify their initial erroneous utterances or whether they report noticing CF. However, regarding modified output, the question that arises is whether it provides evidence of interlanguage development. Although many studies have assessed the effectiveness of recasts by relying solely on modified output, it has rightly been argued that equating it with noticing and L2 development is questionable.

On the one hand, a lack of modified output should not be interpreted as a lack of noticing. Several researchers point out that modified output should be viewed as a discourse phenomenon that may or may not provide insights into psycholinguistic processes associated with noticing. Modified output is voluntary and although the recipients of feedback might notice the target-like construction, they might not necessarily correct their errors (Ellis et al., 2001; Loewen, 2004; Long, 2007; Ohta, 2000). For example, Ohta (2000) found that recasts led to output modification, but not among those learners who received feedback but rather as part of their peers' private speech, defined as "oral language addressed by the student to himself or herself" (p. 52).

On the other hand, it has also been argued that the production of modified output does not necessarily entail deeper processing of the target feature and L2 gains. Gass (2003) explains that when modified output involves mere repetition of a target-like feature, it does not provide robust evidence of L2 development as learners might be “mimicking” their interlocutor without understanding the feature (p. 236). For example, in McDonough and Mackey (2006), learners’ repetitions of target-like questions supplied by recasts were not predictors of question development. In contrast, modified output that involved application of the target model in a novel question was a better demonstrator of L2 improvement. In other words, the ability of the learners to use the target feature in a new linguistic construction during treatment was a better indicator of them having noticed a gap between their erroneous utterances and the target-like construction of recasts.

A few studies have attempted to directly illuminate the relationship between modified output and either the amount of noticing learners reported or the development they demonstrated from a pretest session to a posttest one. Some of these studies show that learners exhibited L2 gains regardless of the amount of modified output they produced (Loewen & Philp, 2006; Mackey & Philp, 1998), whereas others found a relationship between modified output and the amount of noticing reported by learners or their subsequent development (Egi, 2010; Loewen, 2005). A possible explanation for these contradictory findings could be that factors not examined by the researchers above may influence the relationship between modified output and L2 learning. For example, Révész, Sachs, and Mackey (2011) found that the cognitive demands of tasks that learners engaged in affected whether modified output was a predictor of their interlanguage development (see section 2.3 on task complexity). It could tentatively be argued that, apart from task demands, other factors may also impact whether output modification is an indicator of development (e.g. learner factors such as their stage of development, level of proficiency and age, contextual factors such as educational setting and interactional context etc.). It seems that more research is needed to elucidate the role of such variables.

As modified output does not seem to be a robust indicator of L2 outcomes, more reliable measures are required to examine the differential effects of various types of feedback on second language learning. To this end, several empirical studies have employed a pre-test-post-test design in order to gauge L2 development after the provision of feedback. These studies are discussed in the following section.

2.1.5 Studies on Corrective Feedback and L2 Development

Studies on CF and L2 development can be distinguished into (1) those conducted in a laboratory and (2) those conducted in a classroom setting. These studies are discussed in detail in this section and they are also presented in Table 1.

Laboratory vs Classroom Studies

Overall, recasts delivered in laboratories appear to be successful in promoting noticing and subsequent L2 benefits. For example, in Han (2002), a recast group exhibited greater L2 gains related to tense consistency than a control group. Similarly, other laboratory studies have demonstrated that recasts were as effective as clarification requests in facilitating knowledge of the past simple (McDonough, 2007) and knowledge of French grammatical gender (Lyster & Izquierdo, 2009). Recasts also had a similar effect as metalinguistic feedback in the development of dative verbs (Kim & Mathes, 2001) and in the development of the English that-trace filter (e.g. “Who do you think (that) likes ice-cream?”) when measured by a grammaticality judgement test in Goo (2012, 2016) and a written production test in Goo (2012). However, although both types of feedback were more effective than the control condition, metalinguistic feedback was found to be more beneficial than recasts on an oral production test in Goo (2016). It should also be noted that in Goo’s (2012) quasi-experimental study, unlike metalinguistic feedback, the benefits of recasts were mediated by learners’ working memory; however, this was not indicated in Goo (2016) (see section 2.4).

The success of recasts in laboratory studies has been attributed to several factors. For example, most of the time they are delivered during dyadic interactions, are intensive, focus consistently on one error and are also supplied in a context whereby the learners receive individual attention (Han, 2002; Nicholas et al., 2001). Consequently, although recasts are theoretically considered to be an implicit intervention, their provision in laboratory conditions may transform them into more explicit feedback. Probably due to their explicitness, they seem to be more effective in drawing learners' attention to target linguistic areas.

However, several researchers have argued that the benefits of recasts in laboratory studies may not transfer to natural classroom settings. For example, although a group of classroom studies found that receiving CF in the form of either recasts or prompts led to greater L2 development than a condition where no CF was provided (i.e. a control group), when recasts were compared to prompts, the students in the prompts condition exhibited greater L2 gains than the recast groups on both immediate and delayed posttests (Ammar & Spada, 2006; Havranek, 2002; Lyster, 2004; Yang & Lyster, 2010). Nonetheless, as explained earlier, self-repairs may promote the development of linguistic constructions that learners have already partially mastered and to some extent internalized. In contrast, recasts may be an advantageous CF move when they are delivered in the first developmental stages of second language acquisition. For example, Iwashita (2003) found that recasts promoted the development of the Japanese te-form verb, even for learners with very limited prior knowledge of this target feature.

Overall, recasts seem to benefit learners since, in most studies of recasts, recast groups improved their scores significantly more than control groups. Nevertheless, both laboratory and classroom studies have pinpointed several factors that seem to impact on the effectiveness of recasts, such as learners' levels of proficiency, type of target construction and the characteristics of recasts. These factors are discussed below.

Factors Affecting the Effectiveness of Recasts

Level of proficiency

With regard to learners' proficiency levels, previous research has demonstrated that more proficient students reported more noticing and/or exhibited greater L2 development after receiving recasts than less proficient students (Ammar & Spada, 2006; Mackey & Philp, 1998; Philp, 2003; Trofimovich et al., 2007). This finding might be related to Sato's (2011) suggestion that explicit knowledge facilitates the noticing of recasts. That is, learners at higher proficiency levels may have processed the linguistic target of recasts more successfully than those in the first stages of L2 development as a result of having been exposed to the L2 to a greater degree.

Type of linguistic construction

Another factor that appears to affect the usefulness of recasts is the type of target construction. For example, in Leeman (2003,) recasts were more successful when the target error was Spanish number agreement in comparison to gender agreement. Iwashita (2003) also found that recasts were more effective when they addressed the Japanese te-verb form rather than when they targeted Japanese locative constructions. Similalry, Ishida (2004) showed that participants demonstrated greater benefits in the resultative use of the Japanese -te i-(ru) construction in comparison to the progressive use of the same construction after receiving recasts.

These differences in the efficacy of recasts may be related to the salience of the target feature (Ellis, 2007; Yang & Lyster, 2010; Yilmaz, 2012). Long (2007) explains that nonsalient constructions may not be amenable to implicit CF moves, such as recasts, and hence more explicit interventions may be needed. Long's argument is supported by the findings of empirical studies. For example, Ellis (2007) demonstrated that recasts were not more advantageous than a control condition when the linguistic target was the regular past tense -ed, i.e. a non-salient bound morpheme. Similarly, Yang and Lyster (2010) showed that recasts of the regular past simple were less successful than prompts in facilitating L2 development, whereas both types of feedback were equally beneficial when the target construction was a salient irregular past tense. In a similar vein, Yilmaz (2012) exhibited that recasting led to greater L2 gains when the linguistic target was a salient construction. In Yilmaz's study, the Turkish plural was more salient than the Turkish locative in terms of perceptual salience, morphological regularity and similarities between participants' first language (L1) and the L2.

Apart from salience, the degree of structural complexity of a target linguistic construction might also differentially impact on the effectiveness of recasts. For example, in Kartchava and Ammar (2014), recasts targeting English questions were less successful than those addressing the past simple. The researchers explain that, unlike the past tense, the questions involve subject-verb inversion, and hence they are more complex constructions. Likewise, in Long, Inagaki, and Ortega (1998), recasts facilitated the learning of Spanish adverb placement, but not of direct object topicalization, which is a more complex construction. In Ellis (2007), metalinguistic explanations prompting self-repairs benefited learners more than recasts when the target feature was comparative -er, which was coded as grammatically difficult by the researcher. In a similar vein, Guchte, Braaksma, Rijlaarsdam, and Bimmel (2015) found that in a German as a foreign language context (GFL), recasts were less beneficial when the target feature was a complex construction (i.e. the dative case after a preposition), whereas they were as effective as other feedback moves when they targeted a simple construction (i.e. German comparatives).

Characteristics of recasting – degree of explicitness of recasts

Another parameter that might be related to the effectiveness of recasts is their characteristics and degree of explicitness. For example, Loewen and Philp (2006) examined whether the characteristics of recasts described above influenced their ability to induce L2 gains (see section 2.1.3). The participants received CF in a classroom setting and they were administered tailor-made test items based on focus-on-form episodes observed in the classroom. In their study, half of the focus-on-form episodes involved a recast. Other interventions were the provision of explicit information about target errors and elicitations. Loewen and Philp showed that there were no significant differences in the efficacy of the three types of feedback, though the effectiveness of recasts was moderated by their characteristics. In particular, recasts were more beneficial when they were (1) interrogative with rising intonation, (2) short consisting of fewer than five morphemes and (3) one -change recasts. On an explicit-implicit continuum, two of these characteristics (short and one change) may constitute recasts that are a more explicit feedback move. Hence, it could tentatively be argued that more explicit recasts are more successful in drawing learners' attention to linguistic problems and in leading to interlanguage development.

In a similar vein, Nassaji (2009) examined the effectiveness of recasts and elicitations in promoting L2 development of linguistic features arising incidentally during interaction. In order to assess learners' prior knowledge of different linguistic constructions, the participants had to carry out a writing task prior to interaction. In particular, they had to figure out the correct sequence of events depicted in unordered pictures in order to write a story. When the written story was complete, the learners were asked to engage in interaction with a teacher and narrate the story orally. During their dyadic interaction with the teacher, they received recasts and elicitations in a more explicit or implicit manner. After the interaction, the participants performed an immediate postinteraction task that required them to identify the errors in their written descriptions completed prior to the interaction. Two weeks later, they were asked to carry out a delayed postinteraction task that also involved error correction of their first written descriptions. Nassaji examined the effectiveness of explicit and implicit recasts and elicitations by analysing the extent to which the learners corrected their errors in their written descriptions during the immediate and delayed postinteraction tasks. Of the errors corrected, only those that had received feedback during the oral interaction were included in the analysis. The study revealed that in the first postinteraction task, learners were more likely to correct errors addressed by recasts in comparison to those targeted by elicitations; however, when immediate correction rates were compared with those in the delayed postinteraction task, there was a higher decrease in the correction rate for recasts in comparison to elicitations. In other words, learners were more likely to remember linguistic areas addressed by elicitations rather than recasts. Interestingly, in both the immediate and delayed postinteraction tasks, learners were more successful in correcting errors addressed by explicit recasts and elicitations rather than implicit ones.

Previous research has also demonstrated that recasts seem to be more effective when they are combined with other types of feedback, probably because other feedback moves increase their explicitness and make their corrective intention less ambiguous. For example, in Doughty and Varela's (1998) classroom study, when past time references were produced erroneously, recasts were delivered in combination with repetitions of the non-target-like feature. The learners also received both oral and written recasts. The study found that the experimental group exhibited greater L2 gains than the control group on both an immediate and a delayed post-test. Similarly, in Muranoi's (2000) study, recasts of the English article system were successful in inducing L2 learning when they were accompanied by other feedback moves and a more explicit focus on the target construction. In particular, in Muranoi, two experimental groups (A and B) received recasts that were preceded by requests for repetition. However, for group A, the delivery of CF was followed by a formal debriefing focusing on the target construction, whereas for group B, CF was followed by a meaning-focused debriefing (i.e. teachers' comments on the success of conveying messages rather than on accuracy of the target feature). Group C served as a control group that received no CF, only a meaning-focused debriefing. Muranoi found that the greatest improvement was achieved in the explicit condition (i.e. requests for repetition + recasts followed by a form-focused debriefing). This condition was probably more beneficial than the other two as it made the corrective intention of recasts and the target linguistic problem less vague to the students.

The table below provides more details about examples of laboratory and classroom studies investigating recasts. Studies that have examined recasts and one of the main variables of the current empirical work (i.e. mode of interaction, task complexity and individual differences in cognitive abilities such as L2 aptitude) are discussed thoroughly in sections 2.2, 2.3 and 2.4, respectively.

TABLE 1 EXAMPLES OF LABORATORY AND CLASSROOM STUDIES OF RECASTS AND L2 DEVELOPMENT

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
Ammar & Spada, (2006)	Do learners who are exposed to communicative activities with corrective feedback benefit more than those who are exposed to communicative activities only? Are prompts more effective than recasts in leading to L2 development? Are prompts more effective than recasts for both low- and high-proficiency learners?	64 francophone learners of English. They were in grade 6.	Third-person possessive determiners *his* and *her*	Two experimental groups receiving recasts and prompts during communicative activities, and a control group. Prior to feedback, all groups received instruction on the target feature. The participants were administered a pretest, an immediate posttest and four weeks later a delayed posttest comprising passage correction, picture description; the delayed posttest comprised a listening comprehension test and a vocabulary test.	Recasts and prompts were equally beneficial for high proficiency learners. Prompts were more effective than recasts for low proficiency learners.
Doughty & Varela (1998)	Can learners' attention be drawn to formal features without distracting them from their original communicative intent? If so, how? Can focus on form contribute effectively to classroom acquisition?	34 middle-school students from two science classes at an intermediate ESL level	Past time references (simple past and conditional)	Pretest, posttest and delayed post-test design; learners in the experimental group were provided with recasts and repetitions of their errors during three treatment sessions involving oral and written lab reports.	Development by the experimental group, no improvement for the control group. Progress was maintained in an oral posttest but not in a written posttest.
Ellis (2007)	Do recasts have a differential effect on the acquisition of English past	34 low-intermediate adult learners of English	Past simple ("-ed") and comparatives	Conditions: (a) recasts, (b) metalinguistic feedback and (c) control group.	Metalinguistic explanation prompting self-repair led to

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
	tense "-ed" and comparative? Does metalinguistic feedback have a differential effect on the acquisition of English past tense "-ed" and comparative "-er"? To what extent does the effect of corrective feedback on the different grammatical structures differ according to type of feedback?	from a language school in New Zealand.		Feedback was delivered during communicative tasks. L2 development was gauged by a pretest-immediate posttest-delayed posttest composed of an oral imitation test, an untimed grammaticality judgement test (GJT) and a metalinguistic knowledge test.	greater gains than the control condition. However, the gains were immediate for the comparatives and delayed for the past tense "-ed". Recasts had the same effect as the control condition for both constructions.
Goo (2012)	Is there any difference between recasts and metalinguistic feedback in their effectiveness in the acquisition of the English that-trace filter? Does WMC mediate the efficacy of recasts and/or metalinguistic feedback on the acquisition of the English that-trace filter?	95 English as a foreign language (EFL) learners at a Korean university.	The English that -trace filter.	Conditions: (a) recasts, (b) metalinguistic feedback and (c) control group. They performed two one-way information-gap activities. L2 development was measured by a pretest-posttest composed of a GJT and a written production test. WMC was measured by a reading span task and an operation span task.	Recasts and metalinguistic feedback were equally successful in promoting acquisition of the target construction. WMC mediated the effects of recasts but not of metalinguistic feedback.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
Goo (2016)	Is there any difference between recasts and metalinguistic feedback in their efficacy on the acquisition of a L2 syntactic feature? Is the effectiveness of corrective feedback mediated by individual differences in WMC?	83 English as a foreign language (EFL) university learners or elementary school teachers in Korea.	English that-trace filter.	Conditions: (a) recasts, (b) metalinguistic feedback and (c) control group. They performed a dyadic interactional task. L2 development was gauged by a pretest-posttest composed of a GJT and an oral production test. WMC was measured by an operation span task.	Recasts and metalinguistic feedback were equally effective in facilitating L2 development measured by a GJT. Metalinguistic feedback was more effective than recasts in promoting L2 development in an oral production test; however, recasts were also beneficial, as they were significantly more effective than the control condition. WMC mediated the effects of neither recasts nor metalinguistic feedback.
Guchte, Braaksma, Rijlaarsdam, & Bimmel (2015)	Do recasts have a positive effect on the accuracy of new grammar structures? Do prompts, operationalized as metalinguistic feedback followed by elicitation, have a positive effect on the accuracy of new grammar structures? Which type of CF is more effective? Does the effectiveness depend on the targeted structure? Does a student's focus on accuracy have a negative effect on oral fluency?	64 14-year-old 9th grade students (low intermediate) learning German as a foreign language	Dative case after a preposition + comparatives.	Conditions: (a) prompts, (b) recasts and (c) control group. Feedback was given while performing communicative tasks. L2 gains were measured by a pretest, immediate posttest, delayed posttest composed of (a) written accuracy tests involving fill-in-the gap sentences, (b) oral accuracy tests with communicative tasks similar to those used during the treatment and (c) an oral fluency test.	Prompts and recasts led to more gains than the no feedback condition (control). Prompts were more effective than recasts Recasts were more successful when they addressed comparatives than the dative case for written accuracy.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
Han (2002)	Do learners who have received recasts on their L2 output have a greater ability to maintain tense consistency in their L2 narration than learners who have not? Do learners who have received recasts on their L2 output show higher awareness of tense consistency than learners who have not?	8 adult L2 learners of English.	Tense consistency	The participants were assigned to a recast and a nonrecast group. Both groups took part in 8 sessions. L2 gains were measured by a pretest, posttest and delayed posttest administered a month later. The treatment and the tests were composed of written/oral narratives.	Recasts were effective in leading to improvement.
Ishida (2004)	How does overall accuracy in the use of the target feature change? Are the changes related to the total number of recasts for each learner? How does accuracy in the use of the target feature with verbs of different lexical aspect classes/for different aspectual meanings change over time? Are the changes related to the total number of recasts? Are effects of recasting treatment durable?	4 students learning Japanese. 3 of them were L1 English speakers and one was an English-Chinese bilingual speaker.	The Japanese aspectual form -te i-(ru)	The learners participated in 8 conversational sessions. Recasts were delivered during the middle 4 sessions. Two learners participated in a delayed session 7 weeks later, serving as a delayed posttest.	Recasts led to improved performance. The progressive use of -te i-(ru) was less accurate than the resultative use.
Iwashita (2003)	How do NSs respond to NNSs' nontargetlike utterances? Does task-based interaction promote short-term development of grammatical competence among beginning-	55 beginners of Japanese in an Australian tertiary institution. They were native speakers of English or English-Chinese bilinguals.	The Japanese locative, initial construction and a verb morpheme (te-form verb)	The participants were assigned to either a treatment or a control group. The experimental group received implicit negative feedback in the form of recasts and negotiation moves (e.g, confirmation checks, clarification requests),	Positive evidence about the target features was 10 times more frequent than implicit negative feedback. Only learners who had partially mastered the target

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
	level learners of Japanese as a foreign language? What is the relative impact of five types of interactional moves (including implicit negative feedback and positive evidence) on the short-term development of target grammar structures?			and positive evidence (models) while performing three communication tasks (a two-way spot-the-difference jigsaw task, and two one-way information-gap tasks). The study adopted a pretest-immediate posttest-delayed postest design. The delayed posttest was conducted a week later. All of the tests involved oral production (picture description).	constructions benefited from positive evidence. The benefits of implicit negative feedback were not influenced by the participants' knowledge of the target feature prior to the study. Recasts had a larger effect than other conversational moves on short-term L2 grammatical development.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
Kartchava & Ammar (2014)	The research questions were related to the noticeability and effectiveness of three corrective feedback techniques (recasts, prompts and a combination of the two) delivered in the language classroom.	99 francophone college students learning English (high beginners).	Question formation in the past and the past tense	Conditions: (a) recasts, (b) prompts (i.e. repetition, elicitation, metalinguistic information), (c) mixed (both recasts and prompts. Feedback was given to the students during 120-minute sessions while engaging in a communicative task that elicited the two linguistic targets. Noticing measures: Immediate recall protocols. Pretest-posttest: spot-the-differences and picture description tasks.	The recast group reported significantly less noticing than the prompt and mixed groups. No significant differences were found between the prompt and mixed groups. The recast group reported less noticing for the past tense in comparison to the prompt and mixed groups. The recast group reported less noticing for questions than the mixed group. Overall, the past tense was noticed significantly more than questions. The participants exhibited greater development when the target feature was the past tense in comparison to questions. No differences were found among treatment groups for either of the linguistic targets.
Kim & Mathes (2001)	Which form of feedback is more effective for error correction of second language learners, explicit feedback in the form of metalinguistic information or implicit feedback in the form of recasts?	20 Korean adult ESL learners (high, beginners and low intermediate)	English dative alternation	Conditions: (a) explicit metalinguistic feedback, (b) recasts. Controlled production tasks were employed throughout the treatment (two sessions) and tests.	No significant difference in the development of the two groups.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
Leeman (2003)	Does exposure to input with recasts/ with negative evidence/ with enhanced salience of positive evidence lead to greater L2 development than exposure to input with unenhanced positive evidence? If exposure to recasts leads to greater development than exposure to input with unenhanced positive evidence, can these benefits be attributed to either negative evidence or enhanced salience of positive evidence?	74 university students, L2 learners of Spanish, L1 speakers of English	Spanish noun adjective agreement	Conditions: (a) recasts (i.e, negative evidence and enhanced salience of positive evidence), (b) negative evidence, (c) enhanced salience of positive evidence and (d) unenhanced positive evidence (control). The treatment involved information-gap activities. The study utilized a pretest/ immediate posttest/ second-posttest design composed of oral picture-difference tasks. The second posttest was given a week after the treatment.	Only the recast and enhanced-salience conditions significantly outperformed the control group on a posttest. Recasts were more effective when they addressed errors related to Spanish number agreement in comparison to gender agreement.
Long, Inagaki & Ortega (1998)	What is the contribution, if any, of models and recasts in foreign language development?	Study 1: 24 young adult learners of Japanese in their second semester. Study 2: 30 adult learners of Spanish in their third semester.	Study 1: adjective ordering and a locative construction. Study 2: direct object topicalization and adverb placement.	Conditions of both studies: (a) recasts, (b) models, (c) control. Treatment: Study 1: Communication game Study 2: Two communicative tasks	Study 1: Both experimental groups exhibited small improvements in both target constructions. Study 2: No L2 gains were exhibited for direct object topicalization. Recasts led to improvement in adverb placement, while models had no effect.
Lyster (2004)	Will form-focused instruction (FFI) improve French immersion students' ability to accurately assign grammatical gender? Is FFI more effective with feedback than without feedback? Which type of feedback is more effective in FFI—recasts or prompts?	179 10–11-year-old students from immersion French classrooms. They were in fifth grade.	Grammatical gender in French	After receiving form-focused instruction on the target feature, the participants were assigned to three groups receiving recasts, prompts or no feedback. The feedback was provided in six classrooms for 8–10 hours over 5 weeks. A pretest-immediate posttest-delayed posttest was utilized consisting of two written tasks (a binary-choice test and a text-completion test) and	Form-focused instruction was found to be more effective when it was accompanied by prompts rather than recasts or no feedback in both an immediate and a delayed posttest.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
				two oral tasks (an object-identification test and a picture-description test).	
Lyster & Izquierdo (2009)	In the context of dyadic interaction with a researcher, do adult L2 learners of French improve their accuracy in assigning grammatical gender more from prompts or from recasts?	Twenty-five university students, second language learners of French and native speakers of English or bilinguals.	Grammatical gender in French	Conditions: recasts and prompts. A pretest, an immediate posttest and 3 weeks later a delayed posttest were utilized comprising two oral production tasks (an object identification task and a picture description task) and a computerized reaction-time binary-choice test.	Both experimental groups achieved significant gains regarding accuracy and better reaction-time scores, irrespective of feedback type.
Mackey & Philp (1998)	Do learners who participate in task-based interaction with intensive recasts show an increase in developmentally more advanced structures? What is the role of learners' response to recasts?	35 adult ESL learners at beginner to lower intermediate level in Sydney, Australia	Question development	The study had two interactor groups, two recast groups and one control group. Treatment tasks: a picture drawing task, a story completion task and a story sequencing task. L2 gains were measured by a pretest-posttest-delayed posttest (3 weeks later) involving a picture difference task.	In the recast condition, developmentally ready learners exhibited greater gains regarding the use of more advanced constructions in comparison to the interaction group. Recasts promoted short-term benefits.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
McDonough & Mackey (2006)	What is the relationship between recasts, learners' responses to recasts, and their development of ESL question formation?	58 Thai university students, learners of English	Question development	The participants received recasts while engaging in information-exchange and information-gap activities with native English speakers. Each participant was administered an oral production pretest in week 1, engaged in three treatment sessions in weeks 2 and 3, and was administered four oral production tests in weeks 3, 6 and 9.	Recasts predicted question development. Recasts followed by primed production were a predictor of subsequent development, while immediate repetition of recasts was not linked to improvement.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
Muranoi (2000)	Does interaction enhancement (IE), in which a teacher provides implicit negative feedback during an interactive problem-solving task, affect EFL learners' restructuring of their interlanguage article systems? Do two types of IE, differing in the manner of focus on form (implicit negative feedback during an interactive problem-solving task plus formal debriefing versus implicit negative feedback during an interactive problem-solving task plus meaning-focused debriefing), have different effects on EFL learners' restructuring of their interlanguage article systems? If IE, in which a teacher provides implicit negative feedback during an interactive problem-solving task, has an effect on EFL learners' restructuring of their interlanguage article systems, will the effect hold over the five-week post-test period?	114 first-year Japanese college students	English articles	Conditions: (a) IE in the form of requests for repetition and recasts + a formal debriefing involving explicit grammar explanation, (b) IE + meaning-focused debriefing, (c) control, no IE, only a meaning-focused debriefing. The participants were administered a pretest, an immediate posttest and a delayed posttest given five weeks after the treatment. The outcome measures were a GJT, an oral production task and a written production task.	Both experimental groups demonstrated greater gains than the control group in both posttests. IE combined with a formal debriefing had a greater effect than IE followed by a meaning-focused debriefing.
McDonough (2007)	Do clarification requests and recasts differentially impact the emergence of simple past activity verbs?	106 Thai university students, learners of English	The past simple	Conditions: (a) recasts, (b) clarification requests and (c) control group. They performed two-way information exchange and information gap tasks. Each participant was administered an oral production pretest in week 1, engaged in three treatment sessions during	Recasts were as successful as clarification requests in promoting development of the past simple.

Study	Research Question(s)	Participants	Linguistic focus	Design and materials	Results
				week 2, and was administered oral production posttests in weeks 3, 6 and 9.	
Yang & Lyster (2010)	Do the groups that performed form-focused production activities while receiving CF show an overall superiority in learning regular and irregular past tense over the control group, which performed the same communicative classroom activities but without receiving CF on past tense errors? Do recasts have differential effects on the acquisition of regular versus irregular English past tense forms? Do prompts have differential effects on the acquisition of regular versus irregular English past tense forms?	72 EFL students at a university in northern China.	The past simple	Conditions: (a) recasts, (b) prompts and (c) control. Four treatment activities: two dictogloss activities, one question-and-answer activity, and one picture-cued narrative activity. L2 development was gauged by a pretest, an immediate posttest and a delayed posttest administered two weeks after the treatment. The tests consisted of an oral production measure and a written production measure in the form of narratives.	The effects of prompts and recasts were similar regarding irregular past tense forms; however, the prompts were more effective than recasts in leading to accuracy in regular past tense forms.

2.1.6 Summary

In summary, previous research indicates that recasts have the potential to promote noticing and L2 development, especially when their corrective intention is clear to learners. For example, recasts appear to be beneficial when they are provided in a persistent manner, as in laboratory settings, or in more form-oriented classroom conditions. They also seem to be more advantageous when they are supplied in a more explicit manner or in tandem with other feedback moves. Other parameters that influence the effectiveness of recasts are the type of linguistic target and learners' proficiency level. Overall, recasts appear to be less beneficial when they target non-salient, structurally complex constructions, and when they address learners at low proficiency levels in comparison to those at higher levels. Apart from these variables, there have also been calls for research to delve into other factors that might affect the efficacy of recasts, such as mode of interaction, task complexity and L2 aptitude. Although these factors have received some attention in SLA research, their role has not been explored in depth. As they are the focus of the current thesis, they are discussed in detail in the following three sections (2.2, 2.3 and 2.4, respectively).

2.2 Written Synchronous Computer-mediated Communication (SCMC)

An alternative educational context to traditional ELT classrooms is the computer-assisted language learning environment (CALL). From an interactionist perspective, ELT classrooms usually involve face-to-face interaction (FTF), whereas CALL affords computer-mediated communication (CMC). Interaction in the CMC mode is distinguished into asynchronous computer-mediated communication (ACMC) (e.g. by exchanging e-mails) and synchronous computer-mediated communication (SCMC). SCMC is realized as either oral interaction (e.g. via SKYPE) or written interaction in a text-based online chat. The focus of this section is interaction during written SCMC and it is compared with oral interaction in the FTF mode.

2.2.1 Potential Advantages of CMC over FTF Interaction

Previous meta-analyses have indicated that both modes (FTF and CMC) facilitate L2 development (Li, 2010; Lin, Huang, & Liou, 2013; Ziegler, 2015), however, several benefits attributed to CMC make interaction in the online mode a promising learning environment. For example, it has been found that CMC affords greater grammatical accuracy (Salaberry, 2000), the production of more complex constructions (Kern, 1995) and more student-centred interaction, as learners are more active interlocutors who take more turns and initiatives (Chun, 1994; Kern, 1995; Warschauer, 1996). Previous studies have also demonstrated that CMC has a positive impact on several aspects of second language acquisition, such as learners' oral proficiency development (Payne & Whitney, 2002), oral fluency development (Blake, 2009), the acquisition of novel lexical items (Smith, 2004) and the number of learners' self-repairs (Lai & Zhao, 2006).

With regard to CF, its provision during written SCMC has recently witnessed intense interest from SLA researchers due to its interactional similarities with the FTF mode. For example, they both occur in a more informal context than ACMC and entail real-time interaction. In addition, several researchers point out that CF interventions in the SCMC environment have the potential to facilitate noticing of target constructions and subsequent interlanguage development for several reasons (Lai & Zhao, 2006; Sauro, 2009, 2011; Smith, 2012). First, the written nature of the SCMC mode might have a positive influence on noticing and the development of non-salient linguistic constructions. For example, written recasts of bound non-salient morphemes might be effective in increasing their salience and in drawing learners' attention to them. Second, SLA researchers presume that, in comparison to oral recasts provided in the FTF mode, written recasts delivered in the SCMC mode might be more successful in facilitating cognitive comparison between learners' erroneous utterances and the juxtaposed target-like model. The rationale for this presumption is that, during SCMC, students' erroneous output and the target-like model remain permanently on the computer screen. Consequently, they have more time at their disposal to read the input that serves as CF and engage in deeper processing of the linguistic target. Moreover, written recasts can be revised by learners during interaction (Payne & Whitney, 2002; Sauro, 2009; Smith, 2012).

With respect to L2 output, Smith (2012) explains that a positive characteristic of the SCMC mode is that messages are exchanged at a slower pace than in FTF mode and this means that learners can produce and monitor their messages without time pressure. In other words, the SCMC environment may facilitate online planning, i.e. planning one's message during the task (Ellis, 2005b; Yuan & Ellis, 2003). In previous literature, the benefits of online planning have been associated with its potential to free up learners' attentional resources, resulting in greater attention being allocated to form and form-meaning mappings during the formulation of messages (see section 2.3.1). Regarding CF in the SCMC mode, this could imply that after noticing a linguistic problem addressed by feedback, learners might plan the production of a similar construction more meticulously in subsequent messages and avoid similar errors.

Only a handful of studies have attempted to elucidate the extent to which the SCMC mode can promote the noticing of CF and subsequent improved performance. Those that have focused on noticing are presented in the following section, and those that explore L2 development are discussed in section 2.2.3.

2.2.2 Studies on Noticing of Corrective Feedback in Written SCMC

The theoretical rationale for exploring noticing in the FTF and SCMC modes is the Noticing Hypothesis which, as mentioned earlier, posits that noticing linguistic elements in the input is a prerequisite for second language learning (Schmidt, 1990, 1995, 2001) (see section 2.1.2). A study that provides support for Schmidt's hypothesis is Smith (2012). In particular, Smith used stimulated recalls and eye-tracking data to measure the noticing of recasts supplied during written SCMC, and he also employed an immediate and a delayed post-test to gauge L2 development. The study shows that the participants were more likely to exhibit learning of linguistic constructions that they had noticed during interaction.

Considering that noticing could play a facilitative role in subsequent interlanguage development, several SCMC studies have measured the extent to which L2 learners notice computer-delivered feedback by measuring modified output and by employing stimulated recalls, think-aloud protocols and/or eye-tracking data. Empirical research on noticing and CF supplied in the SCMC mode is discussed in greater detail below.

Type of Modified Output

A few studies have shown that learners produce less modified output when they interact in the SCMC mode in comparison to the FTF mode (Gurzynski-Weiss & Baralt, 2014b; Smith, 2010). Due to the low amount of modified output in SCMC and because modified output is not a reliable measure of noticing (see section 2.1.4), Gurzynski-Weiss and Baralt (2014a) used retrospective data in the form of stimulated recalls to examine whether there is a relationship between modified output and the noticing of CF in the two modes. They distinguished two types of modified output: full and partial. *Full modified output* was operationalized as repetition of the whole utterance of the feedback, including the correct feature, whereas, *partial modified output* referred to decoding and producing only the target-like element. Interestingly, the study revealed that, in the FTF environment, both full and partial modified output were indicative of learners' accurate noticing of feedback. In contrast, in the SCMC mode, only partially modified output predicted accurate noticing. The researchers concluded that at least in the SCMC environment, partially modified output is a more robust indication that learners have identified and decoded the target item and engaged in cognitive comparisons between their initial erroneous utterances and the target-like model.

Linguistic Focus

A factor that has been found to impact on noticing is the linguistic focus of computer-delivered feedback. Gurzynski-Weiss and Baralt (2014b) demonstrated that although learners accurately perceived the corrective intention of feedback for the majority of CF episodes in both FTF and SCMC modes, their perceptions were influenced by the types of errors addressed. Specifically, based on participants’ stimulated recall comments, the study showed that, regardless of mode, CF was more effective in facilitating noticing when it targeted lexical errors, followed by semantic errors, and then morphosyntax. The least accurate comments were for errors related to phonology and spelling in the FTF and SCMC modes, respectively.

In line with Gurzynski-Weiss and Baralt (2014b), Smith (2012) also found that the extent to which learners noticed the full recasts provided in written SCMC was associated with their linguistic focus. Smith found that semantic and syntactic recasts were noticed significantly more than morphological recasts, as evidenced in the participants' stimulated recall comments. However, this was not exhibited in eye-tracking data. It should also be noted that, in Smith’s study, the learners had been informed prior to the experiment that their errors would be corrected in the form of recasts. Hence, the recasts were more explicit as the participants were aware of the corrective intention.

Contingency of Recasts

Apart from linguistic focus, another factor that seems to affect the noticing of recasts in written SCMC is their contingency to the target linguistic area. Lai, Fei, and Roots (2008) coded recasts as contingent when they were supplied immediately after learners' erroneous utterances and as non-contingent when discussion or other comments occurred between learners' initial errors and their correction. The researchers measured noticing by requiring that learners think aloud during the interaction and conducting stimulated recalls. The participants reported significantly greater noticing of contingent recasts in comparison to non-contingent ones. Interestingly, the study also indicated that non-contingent recasts were noticed by learners with greater working memory capacity. The researchers drew the conclusion that greater working memory capacity facilitates the noticing of non-contingent recasts as it entails the storage of greater amounts of information and for longer periods of time (see section 2.4.1).

Apart from research on noticing, a few studies have also employed a pre-test, post-test and delayed post-test design to examine the effectiveness of different types of computer-delivered CF, leading to L2 development. As the current section focuses on SCMC, only studies that have investigated CF provided during meaningful interaction will be discussed.

2.2.3 Studies on Corrective Feedback in Written SCMC and L2 Development

While the effectiveness of CF in facilitating L2 development has been investigated to a great extent in FTF mode, only a few studies have explored the effects of written feedback in SCMC mode on subsequent L2 gains (Loewen & Erlam, 2006; Sachs & Suh, 2007; Sauro, 2009; Yilmaz, 2012). These studies are discussed below.

Sachs and Suh (2007) examined the effects of enhanced and unenhanced recasts on the development of reported speech (i.e. backshifting of verbs from past to past perfect). The participants received enhanced or unenhanced recasts during guided story-retelling tasks. L2 improvement was assessed by a pretest and a posttest comprising a paper-based multiple-choice text completion task and a computer-mediated task involving the production of the target feature during interaction. The study demonstrated that recasts promoted development of the target construction regardless of whether they were textually enhanced or not; however, the effectiveness of recasts was not compared with a control group and/or with other types of CF. Only a few SCMC studies have attempted to elucidate this by comparing implicit feedback such as recasts with more explicit interventions such as metalinguistic feedback or explicit correction (Loewen & Erlam, 2006; Sauro, 2009; Yilmaz, 2012).

Loewen and Erlam (2006) investigated the impact of computer-delivered feedback supplied in the form of recasts or metalinguistic feedback on development of the past tense during communicative tasks. The researchers used a pre-test, an immediate post-test and a delayed post-test that included a timed and an untimed grammaticality judgement task to gauge L2 improvement. The study found no advantage of one type of feedback over another and no gains for either recasts or metalinguistic feedback in comparison to a control group.

Sauro (2009) also compared the effects of recasts and metalinguistic feedback during written SCMC. The two experimental groups differed as to whether they received full recasts or metalinguistic information about the target feature, which was the English zero article followed by abstract uncountable nouns, while carrying out collaborative writing activities with a native speaker of English. A computer-delivered acceptability judgement pre-test, immediate post-test and delayed post-test were administered to measure L2 development. Sauro revealed that the metalinguistic group demonstrated significantly greater immediate L2 benefits than the control group; however, there were no significant differences between recasts and metalinguistic information on either an immediate or a delayed posttest.

Interesting insights into the efficacy of mode of interaction and different types of feedback targeting salient vs non-salient features are also provided by Yilmaz (2012). In particular, Yilmaz examined whether there was a difference between (1) explicit correction and recasts and (2) mode of interaction (i.e. written SCMC vs FTF) with respect to their effectiveness in leading to immediate and sustained L2 gains. Yilmaz also explored whether potential L2 benefits of either type of feedback were moderated by (1) mode of interaction and/or (2) salience of linguistic construction. Learners were randomly divided into one of three groups: Explicit correction group, recast group, control group. The explicit intervention involved direct rejection of learners' erroneous utterances and the provision of target-like alternatives which were demonstrated in a direct manner. Recasts were partial, i.e. they only targeted the erroneous parts of learners' utterances. The target items were the Turkish plural and the locative, with the former being more salient than the latter in terms of perceptual salience, morphological regularity and learners' L1-L2 similarity. Thus, the four experimental groups were: SCMC/salient, SCMC/non-salient, FTF/salient and FTF/non-salient. Yilmaz explains that all the participants were placed in two of these conditions and completed a one-way information gap task for each target construction (i.e. one task elicited the plural and the other the locative). L2 learning of the two target constructions was assessed by an immediate and a delayed posttest composed of three tasks: a controlled oral production task that required a two-word description of pictures, a comprehension task that involved verbal prompts requiring learners to select the correct picture, a recognition task (i.e. choosing the correct alternative in order to complete a sentence). A pre-test was not used as the participants were native speakers of English without previous exposure to the L2.

Yilmaz showed that the learners who received either type of feedback outperformed the control group on the oral production test and the comprehension test and only the explicit correction group scored significantly higher than the control group on the recognition test. When the two types of CF were compared, the learners who received explicit correction performed significantly better than the recast group on the oral production and comprehension tests. As for the two learning environments, on the oral production immediate posttest, SCMC led to greater L2 gains than FTF interaction, regardless of feedback type. Similarly, on the immediate and delayed recognition tests, the SCMC condition generated more L2 benefits than the FTF condition. In other words, both explicit correction and recasts were more effective in the SCMC environment on the oral production and recognition tests.

Furthermore, in Yilmaz's study, no relationship was found between type of feedback and mode of interaction. In other words, explicit correction was more effective than recasts, regardless of mode. With respect to the type of linguistic construction, the learners achieved higher scores on the salient construction in comparison to the non-salient one for all outcome measures. No interaction was found between salience of the target feature and type of CF. That is, both explicit correction and recasts were more successful in facilitating L2 learning when they addressed salient constructions.

2.2.4 Summary

In summary, written recasts in SCMC mode were found to be just as beneficial as metalinguistic feedback in Sauro (2009) and in Loewen and Erlam (2006). However, compared to a control group, they were not significantly more effective in generating L2 gains. In Yilmaz (2012), the recast condition was more successful than the control group, but it did not induce significantly greater L2 benefits than the explicit condition. In Yilmaz's study, other factors that influenced the efficacy of recasts were (1) mode of interaction, as written recasts in SCMC mode were more successful than oral ones delivered in FTF mode, and (2) the salience of the linguistic target, i.e. those recasts that addressed salient constructions were more effective than those targeting non-salient linguistic features.

It is also worth noting that the target constructions in Loewen and Erlam (2006) and in Sauro (2009) were also non-salient features: the bound morpheme of the past simple and the English article system, respectively. The low salience of the linguistic targets might have prevented recasts from leading to greater L2 gains than the control groups. Another factor potentially responsible for hampering greater L2 development might be related to the fact that both of these studies employed full recasts. Full recasts reformulate learners' whole utterances without decoding the target linguistic item. For example, Sauro explains that the recasts in her study were quite long and, consequently, despite their written mode, they were highly implicit and less noticeable. Such recasts might be less successful in drawing learners' attention to linguistic features, especially when they are non-salient. With regard to Yilmaz's study, explicit correction was more effective than recasts for both salient and non-salient constructions and in both modes. However, the fact that the participants had no exposure to the L2 prior to the study might be associated with this result. As the learners were not familiar with the linguistic system of the L2, explicit correction may have been more beneficial than more implicit interventions as it drew learners' attention to linguistic items that were part of a completely novel second language.

Although the studies presented above offer invaluable insights into the role of CF in SCMC, more research is needed to elucidate the efficacy of different types of feedback provided during text-based chat and whether their potential benefits are moderated by factors such as educational context, learner characteristics related to age or level of proficiency, individual differences in motivation or cognitive ability and task factors such as task complexity. A study that investigated the relationship between task complexity and mode of interaction is Baralt (2013), which is discussed in the following section devoted to task-based language teaching (TBLT). Studies that examined the effects of computer-delivered pedagogical conditions and individual differences on interlanguage development (e.g. Sagarra, 2007a; Sagarra & Abbuhl, 2013) are presented in section 2.4.

2.3 Task-based Language Teaching (TBLT)

2.3.1 Theoretical Review

As discussed earlier (section 2.1.1), more traditional pedagogical approaches involving synthetic syllabi have received considerable criticism regarding their effectiveness in leading to L2 outcomes. In contrast, analytic syllabi are viewed by SLA researchers as more promising with respect to their potential to facilitate L2 development, especially when they are employed in tandem with a focus on form. Task-based language teaching (TBLT) adopts an analytic approach, whereby the main focus is meaning rather than language; however, it can also be supplemented with a focus on form in order to divert learners' attention from semantic content to certain linguistic constructions when necessary.

A task should satisfy several criteria. According to Ellis (2003), a task is defined as:

> A workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed. To this end, it requires them to give primary attention to meaning and to make use of their own linguistic resources, although

> the design of the task may predispose them to choose particular forms. A task is intended to result in language use that bears a resemblance, direct or indirect, to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills, and also various cognitive processes. (p.16)

One of the most influential TBLT researchers, Long (1985, 2000, 2015) has outlined several key steps involved in designing a task-based syllabus. First, he emphasizes the importance of conducting a needs analysis in order to design target tasks related to what learners do in real-life contexts (e.g. ordering food, making hotel reservations, attending lectures etc.). Second, these target tasks should be classified into different categories and those with common characteristics should fall into the same group. For example, tasks that require learners to attend to different lectures correspond to the same task type and should be in the same category. Third, after different task types have been identified, they should undergo certain adaptations that will make them suitable for learners (e.g. their age, level of proficiency etc.). Learners should perform simple task types at the beginning, and as they improve, they should be presented with more complex task types (e.g. simple dialogues followed by more complex ones). According to Long, in the fourth step, tasks should also be graded and sequenced on the basis of their inherent features that distinguish them in terms of their complexity (see

Robinson's and Skehan's models on task complexity in the following section). Finally, tasks should be implemented and adhere to (1) *methodological principles* related to different areas of research, and (2) those associated with a specific educational setting. For example, interactive and monologic tasks are involved in TBLT methodology; however, which of them will be used in a classroom is a decision made at a local level and may reflect teachers'/ learners' preferences and needs.

From Skehan's (1998) perspective, tasks should be analysed in terms of three aspects: *code complexity, cognitive complexity* and *communicative stress*. *Code complexity* refers to the linguistic complexity of input (e.g. in terms of morphosyntax and lexis). As for *cognitive complexity*, in Skehan's view, its meaning is twofold; first, it is associated with cognitive familiarity (e.g. whether learners are familiar with a topic, discourse genre or task), and second, with cognitive processing (e.g. whether clear and adequate information is provided, how the information is organized etc.). Finally, *communicative stress* is related to factors such as number of interlocutors, time pressure, speed, text length etc. As the current study employed tasks that differ with regard to their cognitive demands (see section 3.4.4), it focuses on Skehan's second factor, i.e. cognitive complexity. The following section introduces the term task complexity from Robinson's and Skehan's perspectives, and it presents two cognitive models that have inspired SLA researchers: the Cognition Hypothesis and the Limited Capacity Model. These models are followed by empirical studies on (1) task complexity and L2 production and (2) task complexity and L2 development. Studies that have explored the effects of task complexity and CF on L2 gains will be discussed in greater detail.

Cognitive Models of Task Complexity

The Cognition Hypothesis

At the outset, *task conditions* and *task difficulty* should be distinguished from *task complexity*. According to Robinson (2001b, 2005, 2011; Robinson & Gilabert, 2007), *task conditions* refer to interactional factors related to learners' participation (e.g. whether a task is open or closed, one-way or two-way) and participant variables (e.g. gender, familiarity etc.). As for *task difficulty*, Robinson argues that the extent to which a task is perceived as difficult is associated with learner factors, which may involve affective variables (e.g. motivation, anxiety, confidence) or ability variables (e.g. aptitude, proficiency, intelligence). *Task complexity* refers to the number of cognitive demands imposed on L2 learners during task performance, irrespective of the learner factors described above. According to Robinson, these demands increase when the task design is manipulated along various dimensions, classified into two categories: *resource-directing* and *resource-dispersing* dimensions (see Table 2).

In Robinson's framework, *resource-directing* dimensions refer to task manipulations that differentiate the cognitive demands of tasks. For example, one such dimension is associated with time references. Simple tasks involve events that take place at the moment of performing the task, in a context shared by all interlocutors (Here-and-Now), whereas complex tasks involve past events that occurred in a distant location (There-and-Then). Another dimension is related to the number of elements (i.e. +/- elements). Specifically, simple tasks have only a few, distinguishable elements, as opposed to cognitively complex tasks that include many, similar elements. A resource-directing dimension in Robinson's model that has received a lot of attention concerns the reasoning demands of a task (+/- reasoning demands). Simple tasks involve mere information transmission, while complex tasks require reasoning about the relationships of different events (e.g. identifying causes and effects of events and providing justifications). When the relationship of events is less obvious, the reasoning demands are expected to be higher. Robinson also distinguishes tasks that require learners to adopt one first-person perspective regarding an event (simple condition) from those that entail many second- or third-person perspectives (complex condition).

The Cognition Hypothesis, formulated by Robinson (2001b, 2003b, 2005, 2011), posits that augmenting task complexity along the *resource-directing* dimensions described above results in more learners' attentional and memory resources being directed to functional and linguistic task demands. Robinson explains that this, in turn, affects both L2 production and development. In particular, he argues that increasing the complexity of monologic tasks in relation to resource-directing variables entails less fluent but more accurate and complex language, and he attributes this to the more complex functional requirements that have to be satisfied during communication. In other words, in Robinson's view, more demanding communicative requirements are expected to induce more attention to L2 speech and, consequently, greater complexity and accuracy of those linguistic constructions needed for the task. For example, regarding linguistic complexity, tasks with greater reasoning demands are likely to lead to the production of more complex syntax and the use of psychological and cognitive state terms (e.g. "believe", "consider" etc.).

With respect to interactive tasks, Robinson presumes that increasing complexity along resource-directing dimensions will bring about more negotiation (e.g. confirmation checks, clarification requests etc.) in tandem with more accurate language, but due to the greater amount of turn-taking, less complex language. He also claims that during interaction under conditions with greater cognitive demands, learners should allocate greater amount of attentional and memory resources to the linguistic constructions embedded in the interlocutor's input. That is, interaction during more complex tasks is expected to encourage learners to attend to linguistic targets of input provided in the form of interactional features such as recasts. Robinson adds that complex tasks are more likely to promote the (1) incorporation of linguistic targets into learners' output (e.g. by producing more uptake) and (2) their long-term retention.

Conversely, according to Robinson's prognosis, increasing the cognitive demands of tasks along *resource-dispersing dimensions* will have differential effects on L2 production. Robinson (2001b, 2005, Robert & Gilabert, 2007) explains that in tasks that are simpler along *resource-dispersing dimensions,* (1) learners are provided with planning time, (2) they have background knowledge (3) they are expected to do only one thing, (4) there is a clear structure, (5) only a single or a few steps are required to complete the task and/or (6) there is no strict sequence of steps. In contrast, under complex conditions, (1) learners have no planning time, (2) they have no prior knowledge, (3) they have to do two or more things simultaneously, (4) there is no clear structure, (5) many steps are involved and/or (6) there is a fixed sequence of steps. For example, if a task requires learners to describe a recipe, it is more complex when there is no planning time and/or the recipe is related to a new unfamiliar cuisine or the learner is not familiar with cooking.

In Robinson's model, greater task complexity related to resource-dispersing factors induces division of memory and attentional resources, resulting in the production of less accurate, less complex and less fluent language during task performance. For example, according to this presumption, L2 learners should produce more errors, syntactically simpler language and more pauses when they have no time to plan their output. It should also be noted that Robinson and Gilabert (2007) warn about possible *synergetic effects* on L2 production when tasks are designed to be more complex along both resource-directing and resource-dispersing dimensions (e.g. tasks with greater reasoning demands without the provision of planning time).

TABLE 2 TASK COMPLEXITY BASED ON ROBINSON'S COGNITIVE MODEL

Simple tasks	Complex tasks
Resource-dispersing variables	
+planning time	-planning time
+structure	-structure
+prior knowledge	-prior knowledge
+single task	-single task
+task structure	-task structure
+few steps	-few steps
+independency of steps (i.e. flexible sequence)	-independency of steps (i.e. strict sequence)
Resource-directing variables	
Here-and-now	There-and-then
-reasoning demands	+reasoning demands
+few elements	-few elements
First-person perspective	Multiple second-/third-person perspectives

Note: Reasoning can be distinguished into spatial, causal and intentional.

The Limited Capacity Model

The Limited Capacity Model questions the predictions of the Cognition Hypothesis with regard to the effects of task complexity on L2 production and development. As explained earlier, the Cognition Hypothesis assumes that when the cognitive demands of tasks increase in relation to resource-directing factors, learners' attention is directed to the linguistic code and, consequently, more accurate and complex language is produced. In contrast, Skehan (2009, 2014b) argues that learners are only equipped with a single pool of limited attentional resources, which might result in trade-off effects during speech production. According to Skehan's Limited Capacity Model, these effects may not allow learners to be fluent and to produce complex and accurate language simultaneously, especially when the cognitive demands of a task increase. Meanwhile, Skehan points out that lower cognitive demands of tasks ease the processing load of learners, and this leads to greater attention to formal aspects of language.

Skehan's explanation of how task complexity affects the allocation of attention across different areas of speech production (i.e. fluency, complexity and accuracy) is inspired by Levelt's (1989) model. Drawing on Levelt's model, Kormos (2011) explains that speech production consists of four stages: *conceptualization* involves planning the content of one's message; *formulation* refers to the grammatical, lexical and phonological encoding of the message; *articulation* is associated with the production of speech sounds; and finally, *self-monitoring*, i.e. evaluating whether the output produced is accurate and appropriate. Kormos explains that for L1 speakers, formulation and articulation are largely automatic; however, for L2 learners, these stages may be influenced by attentional limitations.

Using Levelt's model, Skehan (2009, 2014b) argues that tasks with greater cognitive demands "complexify" the learners' performance at the conceptualizer stage, resulting in (1) fewer attentional resources being devoted to linguistic encoding and (2) the production of less accurate and/or less complex language. With regard to task manipulations, referred to as resource-dispersing dimensions in Robinson's model, such as the provision of planning time, the Limited Capacity Model is in agreement with the Cognition Hypothesis. Drawing on Levelt's model, Skehan contends that when a task provides time to plan the content of a message prior to formulation, it increases the cognitive load and requires greater attentional and memory resources at the conceptualizer stage, whereas it eases the cognitive pressure at the formulation stage. In other words, when learners have already processed the content of the message prior to formulation, they have attentional resources available to allocate to the retrieval of linguistic items during formulation. This might result in the production of more accurate and/or complex language. In contrast, lack of planning time complexifies speech production at the formulation stage, and this may lead to less attention available for linguistic encoding. Likewise, in Skehan's view, greater task complexity in terms of other factors such as lack of structure or prior knowledge (referred to as resource-dispersing in Robinson's framework) is also expected to negatively influence linguistic complexity and accuracy.

With respect to what Robinson calls resource-directing dimensions, Skehan's Limited Capacity Model calls into question the prognosis of the Cognition Hypothesis that greater cognitive demands will direct learners' attention to language and that task complexity will enable them to simultaneously produce more accurate and complex language. But according to the predictions of the Limited Capacity Model, greater reasoning demands, or the manipulation of a greater number of elements, is expected to burden the conceptualizer at the expense of linguistic encoding, and this is likely to result in lower accuracy and/or linguistic complexity. As for the variable related to time references (i.e. there-and-then vs here-and-now), there are contradictory views about which of the two conditions is more cognitively demanding. Counter to the Cognition Hypothesis that considers *there-and-then* tasks to be cognitively more complex than here-and-now ones, Skehan (2009, 2014b) claims that the there-and-then condition might reduce the cognitive demands of a task as learners' encoding of the message at the formulator stage is preceded by processing the input at the conceptualiser stage. In contrast, the here-and-now condition might increase the processing demands for learners, as the linguistic encoding occurs simultaneously with the processing of input. Due to attentional limitations, when there is pressure on both the conceptualiser and the formulator, learners are expected to devote less attention to language during formulation, and consequently to produce less accurate and/or less complex language.

2.3.2. Studies Investigating the Effects of Task Complexity on L2 Production

Many empirical studies have explored the impact of resource-dispersing and resource-directing dimensions on L2 production. With regard to the former, the two cognitive models described above agree that greater task complexity in relation to resource-dispersing factors should negatively influence accuracy and linguistic complexity. Resource-dispersing dimensions that have been explored to a great extent are planning (Ahmadian & Tavakoli, 2011; Crookes, 1989; Foster & Skehan, 1996; Iwashita, McNamara, & Elder, 2001; Kawauchi, 2005; Mehnert, 1998; Ortega, 1999, 2005; Sangarun, 2005; Skehan & Foster, 1997; Tavakoli & Skehan, 2005; Wang, 2014; Wigglesworth, 1997; Yuan & Ellis, 2003), structure (Skehan & Foster, 1999; Skehan & Shum, 2014; Tavakoli & Skehan, 2005; Wang & Skehan, 2014) and prior knowledge/ task repetition (Ahmadian & Tavakoli, 2011; Bygate, 1996, 2001; Gass, Mackey, Alvarez-Torres, & Fernandez-Garcia, 1999; Hawkes, 2012; Kim & Tracy-Ventura, 2013; Lynch & Maclean, 2000, 2001; Mackey, Kanganas, & Oliver, 2007; Pinter, 2005, 2007a, 2007b; Robinson, 2001b; Sample & Michel, 2014; Wang, 2014).

The effects of r*esource-directing* factors on L2 production have also witnessed increased interest from SLA researchers. Previous research has indicated that more complex tasks in relation to resource-directing variables generate more interactional feedback (e.g. clarification requests and confirmation checks) (Gilabert, Baron, & Llanes, 2009[1]; Robinson, 2007b), greater syntactic complexity and more sophisticated vocabulary (Vasylets, Gilabert, & Manchon, 2017), greater lexical variety (Kuiken & Vedder, 2007; Robinson, 2001b) and higher accuracy (Kuiken & Vedder, 2007, 2011; Michel, Kuiken, & Vedder, 2007); however, they have not proved that greater accuracy is accompanied by increased syntactic complexity, as the Cognition Hypothesis holds. In Zalbidea (2017), although more complex tasks led to greater accuracy and syntactic complexity than simple tasks, this difference did not reach significant levels. With regard to time references, both studies inspired by the Cognition Hypothesis (Iwashita et al., 2001; Robinson, 1995) and those following the Limited Capacity Model (Skehan & Foster, 1999; Skehan & Shum, 2014; Wang & Skehan, 2014) have demonstrated that the there-and-then condition facilitates accuracy and/or linguistic complexity to a greater extent than the here-and-now condition[2]. Nonetheless, as explained in the previous section, there are conflicting predictions in Skehan's and Robinson's models as to whether the there-and-then condition is more cognitively demanding in comparison to the here-and-now one.

[1] Gilabert, Baron, and Llanes (2009) found that the complex version of the narrative reconstruction and map task induced a greater number of interactional features, such as clarification requests and confirmation checks; however, this was not observed in a complex decision-making task.

[2] Interestingly, Wang and Skehan (2014) found that greater complexity and accuracy were achieved when learners narrated a video after viewing its content in the there-and-then condition and when the story had a clear structure.

Apart from the impact of task complexity on L2 production, a few studies have also explored whether the effectiveness of CF in assisting interlanguage development is influenced by task complexity. This body of research has compared the effects of feedback supplied during simple and complex tasks on L2 development gauged by pretest-posttest designs. As CF and task complexity are the focus of the current thesis, studies exploring these variables are discussed in detail in the following section.

2.3.3 Studies on the Effects of Corrective Feedback and Task Complexity on L2 Development

A few researchers have explored the influence of CF and task complexity on subsequent L2 development. These studies are distinguished into those that investigate task complexity and interactional features occurring during learner-learner interaction (Kim, 2012; Nuevo, 2006) and the effects of task complexity and CF in the form of recasts supplied to learners by a more advanced interlocutor in the FTF mode (Révész, 2009; Révész, Sachs, & Hama, 2014) or in both the FTF and SCMC modes (Baralt, 2013). These studies are discussed below.

Learner-learner Interaction and Task Complexity

Nuevo (2006) examined the effects of task complexity on learning opportunities and the relationship between task complexity and L2 development of the simple past tense and locative prepositions. One hundred and thirteen adult ESL learners were randomly assigned to one of two experimental groups and a control group. Both experimental groups were asked to carry out two-way interactive tasks in dyads that involved a narrative and a decision-making task manipulated in relation to the +/- reasoning demands variable. In the narrative, the simple and the complex version differed as to whether the pictures were ordered. As for the decision-making task, it asked the learners to decide the best place for people to sit at a party based on information they had about them. The complex condition implicated higher reasoning, as it was more flexible with respect to the decisions the learners could make, while in the simple task, suitable places for each character were more apparent. Learning opportunities during interaction were coded as output modification, self-repair, metalinguistic talk and testing an L2 hypothesis. L2 development was assessed by a pretest, posttest and delayed posttest composed of two oral monologic production tests (a narrative and a description task) and a grammaticality judgement test.

Nuevo's findings regarding learning opportunities were mixed. Simple tasks generated more uptake after recasts, more comprehension checks and other repetitions in comparison to complex tasks. Moreover, in the simple decision-making task, the participants produced more metalinguistic talk than in the complex condition. However, the complex decision-making task led to more confirmation checks than the simple version. Finally, in the narrative task, the learners tested their hypotheses more in the complex version than in the simple one. Thus, the Cognition Hypothesis was only partially supported with respect to its predictions about interactional features and task complexity. As for L2 development, Nuevo did not confirm the Cognition Hypothesis as she found no relationship between task complexity and L2 gains on any of the tests. In other words, engaging in more complex tasks did not lead to greater L2 development than performing simple tasks.

In a similar vein, Kim (2012) explored whether task complexity influenced the number of learning opportunities for question formation during task-based interaction and whether task complexity affects the extent to which task-based instruction facilitates question development in comparison to traditional instruction. One-hundred and one Korean university students were randomly assigned to a control group and three experimental groups and they were required to engage in learner-learner interaction. The three experimental groups differed as to whether they performed simple tasks, +complex tasks or ++complex tasks. Task complexity was addressed by the +/- reasoning demands factor. Specifically, in the simple condition, the learners were asked to exchange information, whereas in the two complex conditions, they had to make a decision; however, the factor that differentiated the +complex condition from the ++complex one was the number of considerations the learners should take into account in order to make a decision. Overall, the simple condition did not involve reasoning, only information transmission, while the + complex condition imposed reasoning demands, but only two considerations should be met, and the ++

complex condition required higher reasoning as four considerations should be taken into account by the participants in order to successfully complete the task.

During these interactions, language opportunities were operationalized as language-related episodes (LREs). Based on Swain and Lapkin (1998), LREs were coded as explicit and implicit feedback or questions about the target feature. Following Pienemann and Johnston's (1987) developmental model of question formation, the researcher measured question development by employing a pretest, posttest 1 and posttest 2. One week after the interaction, the participants were administered a first posttest and two weeks later a second posttest. The tests comprised three tasks: two individual oral production tasks and a paired oral production task. For the former, the first task asked the participants to use the prompts of a story and form questions, and the second task required them to form interview questions. For the paired oral production task, the learners interviewed one another.

With respect to learning opportunities, Kim found that the students in the ++ complex condition produced the highest numbers of question-related LREs, followed by the +complex condition, whereas the simple task elicited the lowest numbers of LREs on questions. As for LREs on developmentally advanced questions (stages 5 and 6 in Pienemann & Johnston's 1987 model), Kim showed that the ++complex tasks generated significantly higher numbers of LREs than the simple task; however, there were no significant differences between the +complex and ++complex tasks or +complex and simple tasks. As for the learners' development on question formation from pretest to posttests, Kim's study indicated that the three experimental groups outperformed the control group that received traditional instruction. Kim explains that all the task-based groups reached a higher level on question formation on the posttest in comparison to the pretest stage. Among the task-based conditions, the ++complex group achieved the highest level, followed by the +complex group, whereas in the simple condition the learners demonstrated the fewest L2 benefits.

Thus, Kim confirmed Robinson's Cognition Hypothesis and she concluded that in comparison to traditional methods, task-based instruction was more successful in promoting question development, especially in the ++complex condition. She attributed the benefits of the more complex tasks to their communicative demands, which might have pushed the learners to produce different types of questions, including more developmentally advanced ones. However, it should be noted that in Robinson's framework, these gains might be exhibited when the target linguistic constructions are relevant to the cognitive-conceptual demands of the tasks, as in question formation; however, it is not clear whether these findings can be generalized to other linguistic elements (e.g. non-salient, communicatively redundant constructions). In other words, the degree to which task complexity facilitates L2 learning might be influenced by the type of linguistic target.

Recasts and Task Complexity during Advanced Interlocutor-Learner Interaction

Révész (2009) investigated the effects of (1) task complexity combined with recasts and (2) task complexity without the provision of recasts on development of the past progressive. Ninety teenagers in an EFL context were randomly assigned to one of four experimental groups and a control group. The experimental groups performed monologic tasks that required them to describe to the researcher a photo depicting what various characters were doing at the time of a crime in order to help a police officer/ researcher find the criminal. Under the complex condition, the learners had to describe the photo without contextual support, i.e. without looking at the photo. Half of these participants received a recast during their description, whereas the other half did not. Under the simple condition, the learners had to describe the same photo with contextual support, i.e. the photo was available during task performance. Similarly, a recast was only provided to half of these participants.

Thus, task complexity was operationalized as the absence or presence of contextual support and the experimental groups were: [+contextual support/photo, +recast], [-contextual support/photo, +recast], [+contextual support/photo, -recast], [-contextual support/photo, -recast]. Nonetheless, as explained earlier, whether the presence or absence of contextual support constitutes a complex task has been the subject of debate. In particular, counter to Robinson’s Cognition Hypothesis that views the presence of contextual support as a simple condition and its absence as a complex condition, Skehan (2009, 2014b) argues that the processing of a stimulus offering contextual support during linguistic encoding may increase the cognitive demands of the task. Hence, Skehan considers the presence of contextual support during task performance as cognitively more demanding than its absence.

With regard to CF, Révész employed simple, isolated, declarative recasts supplied with falling intonation without added emphasis addressing only the target feature. In order to examine L2 benefits, the researcher utilized a pretest, a posttest and a delayed posttest that consisted of three tasks: a written picture description task that indicated whether any gains from the oral feedback had transferred to the written modality, and two oral photo description tasks, one with a photo available during the description and one without.

Révész (2009) found that on the oral description test that had the photo available during the learners' performance, both of the recast groups improved from pretest to te posttest, with the learners in the -photo recast condition scoring higher than those in the +photo recast condition. The non-recast groups showed lower gains and the control group maintained the same score. On the delayed posttest, the +photo recast group and the -photo non-recast group sustained their gains, whereas the -photo recast group had slightly lower scores and the +photo non-recast group demonstrated slight increases. On the oral description posttest that did not have the photo available during its administration, the participants in the -photo recast condition achieved the highest scores, followed by those in the +photo recast condition. As for the non-recast groups, the -photo group scored slightly higher than the +photo group. The control group managed only a slight increase. On the delayed posttest, the +photo recast group maintained its gain as opposed to the -photo recast group, which exhibited a slight decrease. Both of the non-recast groups slightly improved on the delayed posttest, while the control group did not sustain its gains. Finally, on the written description posttest, the learners in the -photo recast condition demonstrated the largest gain, followed by those in the +photo recast condition. The non-recast groups achieved lower gains, with the -photo non-recast group outperforming the +photo non-recast group. The control group had only a small increase. On the delayed posttest, all the groups had lower scores, with the -photo nonrecast group exhibiting the largest decrease, and the -photo recast group the smallest decrease.

Therefore, in Révész (2009), the condition that involved recasts delivered without contextual support was found to facilitate L2 development more than recasts with contextual support or the non-recast conditions. Révész explains that in the +photo condition the learners had to decode both the stimulus (photo) and the feedback (recast). Thus, they had to perform two tasks simultaneously: to process both visual and auditory information. It is possible that the visual stimulus might have prevented the learners devoting sufficient attention to the researcher's recasts. In other words, the photo might have distracted the students from the recasts and, consequently, hampered deeper processing of the target feature. In contrast, in the -photo condition, the participants processed the content of the visual image prior to the delivery of recasts; and consequently, when they received recasts of morphosyntax, their attention was directed to the linguistic target of the feedback. This finding is in line with Skehan's view about contextual support provided in the here-and-now vs. the there-and-then condition and task complexity (see section 2.3.1 above).

Another study that offers interesting insights into the relationship between task complexity and the effectiveness of recasts is Baralt (2013). Baralt also went a step further as she illuminated this relationship for different modes of interaction. The participants were 84 adults learning Spanish as a foreign language and the target construction was the Spanish past subjunctive. The study had four experimental groups and a control group; two of the experimental groups engaged in one-to-one computer-mediated communication (CMC), whereas the other two interacted in the face-to-face (FTF) mode. In each mode, half of the participants performed a cognitively more demanding task, whereas the other half completed a simple version of it. Thus, the experimental groups in the study were: [CMC, -complex], [CMC, +complex], [FTF, -complex] and [FTF, +complex]. In all the experimental conditions, recasts of the target feature were given with falling intonation at the end, without stress in the FTF mode and without added enhancement in the CMC mode.

Task complexity was operationalized as +/- intentional reasons. Specifically, the learners in all groups were provided with stories similar to real-life situations and were required to retell these stories using comic cards. However, what differentiated the complex from the simple version was whether the participants had to think of the characters' intentional reasons by themselves or whether this information was provided by the task. The former was coded by the researcher as more complex than the latter due to its greater reasoning demands. In order to validate the construct of task complexity, Baralt employed a perceived difficulty questionnaire and retrospective time-on-task judgements. The former sought to gauge learners' perceptions with regard to task complexity. Retrospective time-on-task judgements required the learners to estimate how much time they spent on the task. It was expected that, for the complex task, the learners would judge that they had spent a greater amount of time than they had, whereas the opposite should occur for the simple version. With respect to language development, it was assessed by a pretest, an immediate posttest and a delayed posttest administered a week later. The tests consisted of three tasks: two productive tasks, one in FTF mode and one in CMC mode, to show whether any gains were transferred from one mode to the other, and one receptive multiple-choice test. Interestingly, the tests included both non-tailored and tailored items that resulted from the intentional reasons that the learners thought of by themselves in the complex condition. This distinction sought to indicate whether one of the two types of items (tailored or not) would be produced more accurately by the learners.

Regarding the validation of task complexity, Baralt's study found that the mode of interaction affected the learners' perceived difficulty of the task on a post-task perception questionnaire. In particular, when the two modes were compared, Baralt found that the FTF groups judged the tasks to be more difficult than the CMC groups, irrespective of whether they performed the simple or the complex version. However, Baralt reported that the retrospective time-on-task judgements validated task complexity as they indicated that, in the complex condition, the learners in both modes judged that they spent more time than their actual time. As for L2 development, Baralt found that on the three assessments, the CMC -Complex group and the FTF +Complex group achieved greater development than the CMC +Complex, FTF -Complex and control groups. The researcher also found that the FTF +Complex group exhibited the greatest gains on the productive tasks of the delayed posttest as opposed to the CMC +Complex group which had the lowest scores. When the tailored vs non-tailored items were examined, the study indicated that the learners' production of the past subjunctive improved more when the items were tailored on the immediate and delayed production posttests. However, no differences were found between the two types of items on the multiple-choice receptive test.

Baralt argues that her study only supports the Cognition Hypothesis in the FTF mode, whereby the +reasoning demands condition was more effective than the -reasoning demands condition in leading to subsequent L2 benefits. As for the CMC mode, the researcher pointed out that in the +reasoning demands complex condition, where the learners exhibited the least development, there were longer and confusing turns and non-contingent recasts. This might have hindered a cognitive comparison between the erroneous linguistic construction and the target-like model provided by feedback (see section 2.2). In contrast, Baralt adds that deeper processing of the target construction was achieved in the CMC -Complex condition because it involved fewer turns, shorter discourse and recasts delivered immediately after errors.

Nevertheless, the question that arises is whether the condition where the learners had to think of characters' intentions by themselves was more cognitively demanding than the one that involved pre-determined intentions. First, the post-task perception questionnaire did not validate any differences between the two task versions in terms of their cognitive demands. Second, although in the +intentional reasons condition the learners had to think what prompted the characters to do certain actions, the -intentional reasons condition that provided the participants with the characters' intentions may have implicated the processing of more visual elements (i.e. pictures accompanied by a written script of intentions). As explained earlier, previous studies have demonstrated that learners' output is less accurate when they have to process the content of a video while engaging in linguistic encoding in the here-and-now condition (Skehan & Foster, 1999; Skehan & Shum, 2014; Wang & Skehan, 2014) and that they benefit less from recasts on morphosyntax delivered while processing the content of a photo (Révész, 2009). Similarly, Révész, Sachs, and Hama (2014) argue that in Baralt's study the written script of intentional reasons might have required processing of more information and thus increased the cognitive load for the participants during oral interaction in the FTF mode.

Révész, Sachs, and Hama (2014) also delved into whether task complexity, addressed as +/- reasoning demands, impacts on the effectiveness of recasts in promoting interlanguage development. However, unlike the other two studies (Révész, 2009 and Baralt, 2013), the simple and complex tasks did not differ in the numbers of visual stimuli or in the range of input sources that had to be processed while receiving feedback. The target linguistic area was the English past counterfactual construction. The participants carried out two tasks that required them to read a story before being presented with different events whose relationship was cause and effect. By using two clauses (i.e. *if* and *then*), the learners were expected to identify the causes of several events; however, this relationship was more obvious in the simple condition, whereas it posed greater reasoning demands in the complex condition. Other than that, the tasks in the two conditions involved the processing of the same number of visual elements.

The construct of task complexity was validated by three methods. First, two TBLT experts were asked to judge the complexity of the two task conditions. Second, a *dual-task methodology* was applied. This technique required the learners to perform a primary task (i.e. relate causes and effects of events) and simultaneously do a secondary task. The secondary task involved a change of colour on the computer screen (green or red) and the participants had to react as quickly as possible to only one of the two colours. The participants were expected to react more slowly during the complex task due to its greater cognitive demands. Finally, eye-tracking was used, and it was predicted that the complex tasks would involve longer eye-fixations on the visual stimuli of the task. All three methods revealed that the tasks functioned as more complex or simple in the direction designed by the researchers. Regarding L2 development, it was captured by three outcome measures: *an oral production test*, *a meaning-to-form written test* that involved fill-in-the gap items and *a form-to-meaning written test* which consisted of multiple-choice items.

The study showed that task complexity affected the extent to which recasts promoted L2 development but in a different direction from that predicted by the Cognition Hypothesis. Specifically, the students who received recasts while engaging in simple tasks benefited more than those who performed complex tasks. However, this difference was significant only on the oral production test, and not on the two written tests. The researchers' explanation of this finding was twofold. First, task complexity did not generate greater L2 gains due to synergetic effects that may occur when tasks are manipulated simultaneously in relation to resource-directing and resource-dispersing variables (Robinson & Gilabert, 2007). Second, drawing on Skehan's Limited Capacity Model and Levelt's speech production model, the researchers argued that when the learners performed complex tasks, they allocated more attentional resources to the processing of the content of pictures and engaging in macroplanning, and consequently less attention was available for the processing of recasts and their linguistic target. This was reinforced by eye-tracking data that indicated longer eye fixations on pictures in the complex condition.

2.3.4 Summary

In summary, more cognitively complex tasks along resource-directing dimensions may not be effective in drawing learners' attention to language, as the Cognition Hypothesis posits. Previous studies have demonstrated that increasing task complexity does not lead to simultaneously more complex and accurate language during L2 production. Moreover, there is no evidence that more complex tasks promote greater processing of CF and L2 development. Nevertheless, other factors that appear to influence the effectiveness of feedback and the production of accurate language are the number of visual elements, such as photos, videos and written scripts, that have to be processed during the formulation of a message. Based on the Limited Capacity Model, the processing of such elements may increase the cognitive load in the conceptualizer stage, resulting in less attention to grammatical, lexical and phonological encoding in the formulation stage.

Another important variable that may impact on the effectiveness of task complexity and CF in facilitating interlanguage development is individual differences in cognitive abilities such as L2 aptitude. This variable is discussed in the following section.

2.4 Individual Differences in Cognitive Ability

Individual differences refer to several constructs, such as learners' cognitive ability, motivation and anxiety. With regard to cognitive ability, previous research has shown that working memory capacity (WMC) and L2 aptitude play a pivotal role in second language acquisition. As WMC has been found to influence the effectiveness of CF, the following section briefly discusses an influential model of WMC formulated by Baddeley (2003) and its relationship with L2 learning and, specifically, with recasts. However, the main focus of the current thesis is L2 aptitude. First, a theoretical review of L2 aptitude will be presented, followed by studies that have examined the influence of learners' aptitude (1) on ultimate attainment and (2) on the effectiveness of different types of instruction and CF.

2.4.1 Baddeley's Model of Working Memory Capacity (WMC)

Working memory has been defined by Baddeley (2003) as "the temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities" (p. 189). The model was initially developed by Baddeley and Hitch (1974) and later updated by Baddeley

(2000, 2003). Baddeley's model views WM as a multicomponent system that involves a central executive system supported by two storage components: a phonological loop and a visuospatial sketchpad. The former functions as a temporary storage system for verbal and acoustic information, whereas the latter is responsible for storing and processing visuospatial information. Baddeley (2000) added a fourth component, an episodic buffer. Its role is to integrate visual, spatial and verbal information from the phonological loop and visuospatial sketchpad. In Baddeley's model, the phonological loop consists of two subcomponents: a phonological store and a system for articulatory rehearsal. With respect to the phonological store, it can hold verbal information temporarily, while the articulatory rehearsal process is similar to subvocal speech and its role is to retrieve and rearticulate fading verbal information in the phonological store in order to revitalize it (Baddeley, 2003).

Regarding WMC and SLA, Baddeley (2003) contends that "phonological loop capacity is a good predictor of the ability of children and adults to learn a second language" (p. 832). In line with this argument, several studies have shown positive correlations between learners' phonological short-term memory (PSTM) and several aspects of L2 development. For example, a relationship has been demonstrated between PSTM and speech production (Kormos & Sáfár, 2008; O'Brien, Segalowitz, Freed, & Collentine, 2007), the ability to learn new vocabulary (Atkins & Baddeley, 1998; Cheung, 1996; Daneman & Case, 1981; Gupta, 2003; Masoura & Gathercole, 2005; Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995; Speciale, R. Ellis, & Bywater, 2004) and the ability to acquire new grammatical constructions (Daneman & Case, 1981; N. Ellis & Sinclair, 1996; N. Ellis & Schmidt, 1997; O'Brien, Segalowitz, Collentine, & Freed, 2006; Williams & Lovatt, 2003). The central executive also plays a pivotal role in several aspects of SLA. For example, previous research has indicated that it is associated with L2 comprehension (Harrington & Sawyer, 1992; Kormos & Sáfár, 2008; Leeser, 2007; Walter, 2004), production (Kormos & Sáfár, 2008; Zalbidea, 2017), morphosyntactic processing (Juffs, 2004; Sagarra, 2007b) and the acquisition of grammar (Harrington & Sawyer, 1992; Kempe & Brooks, 2008; Leeser, 2007).

Many researchers have also explored the relationship between long-term memory (LTM) and PSTM, and they point out that there is an interaction between them that supports the acquisition of lexis or morphosyntax. As explained earlier, the PSTM serves as a temporary storage and processing system for lexis or morphosyntactic constructions. However, their processing in the phonological loop can be assisted by the LTM, which involves a larger database with previously stored information (Gathercole, Hitch, Service, & Martin, 1997). For example, Gathercole and Baddeley (1993) found that it was more difficult for participants to recall and repeat non-wordlike items that were very different from real lexis than those similar to existing lexis. In addition, Masoura and Gathercole (2005) demonstrated that the amount of participants' knowledge of vocabulary acquired prior to their study (i.e. their lexicon) influenced their speed of learning novel lexical items. These findings suggest that information retrieved from the LTM might facilitate the processing of new information in the PSTM if that new information resembles representations already stored in the LTM. However, it should be noted that although the LTM might assist information processing in the PSTM, the capacity of working memory is not restricted to reactivating old information but can also generate new representations (Baddeley, 2003). In addition, it has been demonstrated that PSTM can also facilitate LTM, and in particular, the consolidation of information regarding lexis or morphosyntax (N. Ellis, 1996; N. Ellis & Schmidt , 1997; N. Ellis & Sinclair, 1996).

2.4.2 WMC and Corrective Feedback

From an interactionist perspective, a few studies have attempted to elucidate the role of WMC in facilitating noticing of CF. Mackey, Philp, Egi, Fujii, and Tatsumi (2002) investigated the extent to which WMC is related to noticing recasts that target errors in question formation, and whether this relationship is associated with learners' developmental level. The study showed that participants with smaller WMC were more likely to report less noticing during stimulated recalls in comparison to those with larger WMC. However, the relationship between noticing and WMC was only marginally significant. The study also indicated that learners with high PSTM reported greater noticing, but mainly at lower developmental stages. Nevertheless, the researchers warn that due to the small number of participants, these findings should be interpreted with caution and more research is needed. In a more recent study, Mackey, Adams, Stafford, and Winke (2010) explored the relationship between learners' WMC and the amount of modified output they produced. The researchers demonstrated a positive relationship between WMC and output modification. That is, learners with larger WMC produced a greater amount of modified output than those who achieved lower scores on a WMC test. Nonetheless, it should be noted that due to the low effect sizes (under 20%), the researchers suggested that output modification might be associated with other variables as well, apart from WM.

Considering that modified output is not a reliable predictor of subsequent improvement (see section 2.1.4), researchers have also utilized a pre-test, post-test and delayed posttest design to illuminate whether the effectiveness of CF in terms of promoting L2 gains is influenced by learners' WMC. Mackey and Sachs (2012) examined the relationship between WMC and L2 development of nine older learners whose ages varied from 65 to 89. The learners received interactional feedback targeting errors in question forms while performing communicative tasks. The researchers found that only the central executive played an important role in immediate L2 benefits. With regard to delayed posttests administered both a week and a month after the treatment, no relationship was found between the learners' improved performance and any of the components of WMC.

In a similar vein, Révész (2012) explored whether PSTM and complex WMC are associated with potential positive effects of recasts on development of the past progressive across different outcome measures. L2 development was gauged by a pre-test and a post-test that consisted of a grammaticality judgement task, a written picture description task and an oral description task. The study revealed that recasts were effective in leading to second language learning; however, complex WMC was linked to L2 benefits on the written tests (i.e. a grammaticality judgement and written picture description test), while greater PSTM was related to improvement on the oral test.

Goo (2012) obtained similar findings to Révész (2012). Goo examined the relationship between WMC and the effectiveness of recasts and metalinguistic feedback in promoting development in the knowledge of the English that-trace filter. L2 development was assessed by two outcome measures: a grammaticality judgement test and a written production test. The study found that although the two types of feedback were equally effective on both tests, the L2 benefits demonstrated by the recast group on the grammaticality judgement test and the written production test were mediated by their WMC. In other words, learners with larger WMC benefitted from recasts to a greater degree than their counterparts with smaller WMC. Hence, in both Révész (2012) and Goo (2012), the central executive affected the extent to which recasts facilitated L2 gains exhibited on written tests. Nevertheless, in a more recent laboratory study that was a conceptual replication of Goo's (2012) quasi-experimental study, Goo (2016) demonstrated that WMC was associated with the benefits of neither metalinguistic feedback nor recasts (see Table 1). Unlike his earlier study, Goo (2016) employed an oral production test rather than a written one, and feedback was provided during a dyadic interactional task. Goo (2016) concluded that providing recasts during dyadic interaction in a laboratory setting benefits learners, irrespective of their WMC.

Yilmaz (2013) also offers interesting insights regarding the relationship of implicit and explicit feedback with WMC. In particular, Yilmaz investigated the efficacy of explicit correction and recasts in assisting the acquisition of two Turkish constructions: the plural morpheme and the locative case morpheme, and the role of WMC. The participants were native speakers of English and L2 learners of Turkish; however, they had no previous exposure to the L2 prior to the study. Learners' emerging knowledge of the target features was assessed by an oral production, a comprehension and a recognition test. A delayed post-test with the same tasks was employed two weeks later. Similar to Goo (2016), Yilmaz (2013) found no relationship between WMC and knowledge of either the plural or the locative under the recast condition. On the contrary, unlike Goo (2012, 2016), in Yilmaz's study, WMC moderated the degree to which the learners benefited from explicit correction targeting either the plural or the locative. However, a comparison between Yilmaz's and Goo's studies should be made with caution. In Yilmaz (2013), the participants had no exposure to the L2 prior to the study, whereas in Goo (2012, 2016) the learners were familiar with the L2. The rationale for why this difference might be crucial is that when an experiment is conducted in a context familiar to the participants' linguistic system, information stored in LTM with respect to the second language (e.g.

moprhosyntax, phonology etc.) may assist in the processing of more implicit feedback in WM. In contrast, the processing demands of completely novel linguistic items in an unfamiliar L2 involve the creation of new representations; consequently, more explicit interventions may be needed for large WMC to contribute, as in Yilmaz (2013).

Previous findings have also indicated a relationship between WMC and sustained L2 gains. In particular, a few studies have shown that the effect of WMC on the efficacy of recasts in facilitating L2 development was revealed in a delayed post-test rather than an immediate one (Mackey et al., 2002; Trofimovich, et al., 2007). In Mackey et al. (2002) the participants received recasts on question formation and their L2 development was measured by a pre-test, an immediate post-test and a delayed post-test composed of communicative tasks. The study demonstrated that students with high WM scores achieved greater L2 development than those with lower WM scores; however, this difference was only found in a delayed posttest. Likewise, Trofimovich et al. (2007) found that their participants' phonological memory influenced the benefits of morphosyntactic recasts, though this relationship was revealed only in a delayed post-test. Therefore, the findings of these studies suggest that WM might not affect L2 learning immediately after treatment but rather in the long run.

Regarding computer-delivered feedback in the form of oral recasts, Sagarra (2007a) noted that learners with higher WM spans produced more modified output and they exhibited greater gains related to accuracy than low WMC learners. Additionally, in a more recent study, Sagarra and Abbuhl (2013) showed that the extent to which computer-delivered recasts are influenced by WMC is associated with whether they are oral or written. In particular, Sagarra and Abbuhl found a positive relationship between WMC and the efficacy of both enhanced and unenhanced computer-delivered oral recasts, while WMC did not affect computer-delivered typographically enhanced and unenhanced written recasts. This finding is in line with Payne and Whitney (2002), who observed that WMC influenced oral FTF interaction to a greater extent in comparison to interaction combining both modes (oral FTF and written SCMC). However, it should be noted that, in Sagarra and Abbuhl (2013), recasts were provided for sentences that were not part of meaningful interaction. Hence, their focus was not on SCMC. In future, more studies are needed to delve into the impact of WMC on feedback supplied as part of meaningful interaction in both SCMC and FTF modes.

With respect to TBLT, an under-researched area is the effects of task complexity on the efficacy of CF and the role of WMC as a moderating variable. A study that has attempted to shed light on this issue is Kim, Payant, and Pearson (2015). The researchers explored the degree to which task complexity with differential reasoning demands affected (1) noticing of recasts on non-targetlike English question constructions, (2) development in the knowledge of a target feature and (3) whether these relationships are influenced by WMC. The study was conducted in a laboratory and required the participants to perform three two-way information-gap tasks with a native speaker. The two experimental groups differed as to whether they carried out simple or complex tasks. In the simple condition, the learners had to exchange information, whereas in the complex condition, they were asked to exchange information, and additionally to make comparisons and evaluate information in order to make a decision and reach a consensus. The learners in both conditions received a recast in response to non-targetlike questions. The construct of task complexity was validated by utilizing stimulated recalls immediately after the third treatment, during which the participants were presented with a video showing the treatment sessions and asked to report their thoughts during the interaction. The noticing of recasts was measured by using immediate cued recalls; inaccurate production of question formation was followed by a recast and the researcher knocking twice, thus prompting the learner to repeat the preceding utterance, i.e. the recast. L2 development was assessed by a pretest, posttest and delayed posttest that consisted of three oral production tasks.

The study found that regardless of task complexity, both groups recalled about 80 per cent of the recasts accurately in immediate cued recalls; however, WMC influenced their ability to notice recasts and repeat them correctly. Specifically, learners with larger WMC noticed feedback to a greater degree than their counterparts. A similar pattern was found for recasts on advanced question forms (stages 4 and 5). In particular, the learners in both conditions (simple and complex) accurately recalled about 78–79 per cent of recasts; however, the learners' WM predicted the extent to which the participants successfully noticed feedback. With respect to the development of question formation, the study demonstrated that learners with larger WMC benefited more from recasts than their low WMC peers, especially under the complex condition.

2.4.3 Summary

In summary, previous research has provided evidence that both the central executive and PSTM influence several areas of second language development, such as the acquisition of vocabulary and grammar, L2 comprehension and production. Previous studies also indicate that WMC is associated with the extent to which different types of CF facilitate the noticing and learning of linguistic targets. Apart from WMC, researchers interested in individual differences in SLA have also examined the role of L2 aptitude in learning outcomes. The following section first presents a theoretical review of L2 aptitude, and second studies that have investigated the relationship of L2 aptitude (1) with learners' ultimate attainment, (2) implicit and explicit instruction and (3) different types of CF.

2.4.4 A Theoretical Review of L2 Aptitude

Language aptitude refers to cognitive and perceptual abilities that facilitate second language acquisition (Carroll, 1965, 1981; Granena, 2013). At the outset, it should be noted that aptitude is a different construct from intelligence and affective variables such as motivation and anxiety. It is related to WMC and it is a predictor of general L2 proficiency and several areas of L2 learning (see the meta-analysis conducted by Li, 2016). According to Carroll's proposals, aptitude involves abilities associated with four components: *phonetic coding ability*, *inductive language learning ability*, *grammatical sensitivity* and *rote learning ability* or *associative memory*. *Phonetic coding* refers to the ability to identify sounds, to make connections between sounds and their symbols, and to retain them. *Inductive language learning* is the ability to induce rules from input. *Grammatical sensitivity* is the ability to identify the functions of words in sentences; and finally, *rote learning* is the ability to identify connections between sounds and meanings, and retain them. Based on Carroll's work on L2 aptitude, the Modern Language Aptitude test (MLAT) was created by Carroll and Sapon (1959). Apart from inductive learning, the MLAT was designed to tap into the other three aptitude components of Carroll's framework (i.e. *phonetic coding*, *grammatical sensitivity* and *associative memory*) by using subtests such as Number Learning, Phonetic Script, Words in Sentences and Paired Associates.

From an information processing perspective, Skehan (1998) argues that aptitude consists of cognitive factors that involve *memory-as-retrieval*, *phonetic coding* and *language analytic ability*, which includes Caroll's *grammatical sensitivity* and *inductive language learning*. An updated model constructed by Skehan (2002) views *attentional control* and *working memory* as important components of L2 aptitude. Considering working memory as an aptitude construct was an important update that has influenced contemporary aptitude research and measures. Interestingly, Skehan (2002, 2016) and Wen, Biedroń, and Skehan (2017) also suggest that different aptitude constructs are associated with the cognitive processes involved in three SLA stages: *language input*, *central processing* and *language output*. Cognitive processes occurring in the first stages of development are *language input processing*, *noticing* and *patterning*. *Noticing* linguistic features entails directing attention to them so that initial representations are created in long-term memory (Schmidt, 2001; Long, 2015). Skehan (2016) explains that noticing is related to input processing; however, it additionally implicates language analysis, at least to some extent. In Skehan's framework, a more advanced stage than noticing is *pattern identification,* which entails attention to wider linguistic constructions such as word order or the relationships between different linguistic items. The initial stage of identifying language patterns is in turn followed by extending, complexification and restructuring them. Skehan (2016) considers these stages as "a leap in the process of acquisition" (p. 22), as they involve deeper understanding of linguistic elements. In line with previous research that has revealed the benefits of feedback (see section 2.1), Skehan argues that all of the above stages (i.e. noticing, pattern identification and extension/ complexification/ restructuring) are reinforced by the provision of CF.

In Skehan's model, the remaining stages are differentiated from those presented above, as they involve learners' control over their output. These stages are also divided into two categories: (1) those associated with gaining control over the L2 system, these involve first, error avoidance during slow and effortful production, and later, *automatization,* i.e. production without consciousness and monitoring, and (2) in the most advanced stages of SLA, these entail *salience/ repertoire creation* and *lexicalisation*. The former refers to the use of lexical items and collocations that reflect native-like performance, and the latter is concerned with the ability of advanced speakers who have reached native-like level of L2 production to retrieve expressions as chunks.

Skehan (2002, 2016) posits that there is a relationship between the cognitive processes described above and different components of L2 aptitude. In particular, in Skehan's framework, the role of *attentional control* and *working memory* is viewed as pivotal during *language input processing*. The contribution of *phonetic coding ability* and *working memory* is also considered crucial for *noticing* and *pattern recognition*. *Pattern recognition*, *complexification* and *handling of feedback* are expected to be linked with both *working memory* and *language analytic ability*. Skehan also relates *error avoidance* to *working memory* and *retrieval memory*; and in the more advanced stages of language output, he argues that (1) *automatization* is expected to be associated with *retrieval memory*, (2) the *creation of a repertoire* with *retrieval memory* and *chunking* and, finally, (3) *lexicalization* with *chunking*. Table 3 below presents Skehan's (2016) aptitude model.

TABLE 3 SKEHAN'S MACRO-SLA APTITUDE MODEL (FROM SKEHAN, 2016)

SLA stages	L2 cognitive processes	Aptitude Constructs
Language input Central processing Language output	Input processing (segmentation)	Attentional control Working memory
	Noticing	Phonetic coding ability Working memory
	Pattern recognition	Phonetic coding ability Working memory Language analytic ability
	Complexification	Language analytic ability Working memory
	Handling feedback	Language analytic ability Working memory
	Error avoidance	Working memory Retrieval memory
	Automatization	Retrieval memory
	Creating a repertoire	Retrieval memory Chunking
	Lexicalization	Chunking

Since the release of MLAT, updated aptitude measures have been developed taking into account the suggestions of more recent work on L2 aptitude. For example, the CANAL-F test designed by Grigorenko, Sternberg, and Ehrman, (2000) has focused more on learners' ability to cope with novel

L2 learning conditions. HiLAB, created by Linck, Hughes, Campbell, Silbert, Tare, Jackson, and Doughty (2013), addresses learners at more advanced levels and puts more emphasis on the role of working memory. In particular, it provides measures related to the functioning of the central executive component of working memory, phonological short-term memory (PSTM), associative memory, long-term memory retrieval, processing speed, implicit learning and auditory discrimination.

Another recent aptitude test that the current study also employs is the LLAMA aptitude test developed by Meara (2005). The LLAMA test is composed of four subtests that serve as measures of rote, associative memory (LLAMA B), the ability to recognize patterns in spoken language (LLAMA D), the ability to associate sounds with symbols (LLAMA E) and inductive language learning ability (LLAMA F). An exploratory validation study conducted by Granena (2013) showed that the participants' L1 (Chinese, Spanish and English) and gender were not related to their scores on LLAMA tests. Moreover, Granena's (2013) study indicated that the LLAMA D component might be a measure of implicit learning, while the other three components might measure explicit learning. In particular, Granena found no correlations, or very weak ones, between LLAMA D and LLAMA B, E and F. That is, there were participants with high scores on the LLAMA D subtest but low scores on the other three components (i.e. B, E and F), and vice versa. Granena explains that a common feature of LLAMA B, E and F is that learners can work out relationships between objects and names in LLAMA B, sounds and symbols in LLAMA E, and grammatical rules in LLAMA F. Although all of these subtests are timed, they involve a study phase that allows learners to hear or view the same features more than once, and employ cognitive ability related to explicit learning. In contrast, in LLAMA D, learners have to listen to oral patterns only once in order to recognize them later. Granena points out that unlike the other three components, LLAMA D does not involve a study phase and does not require learners to utilize their analytical ability in order to figure out relationships between elements. Nonetheless, associating LLAMA D with exclusively implicit learning conditions could be misleading. LLAMA D may also be a measure of learners' ability to consciously register oral patterns, store them in their phonological WM, and subsequently recognize them when exposed to them. Although learners would not use their analytical ability, this process would probably entail the use of cognitive ability related to explicit learning. However, further research is needed to illuminate the role of LLAMA D in implicit and/or explicit learning conditions.

Regarding existing aptitude research, many studies have investigated aptitude in two ways: (1) as a variable that can predict learners' achievements in the L2 or their ultimate attainment and (2) as a variable whose influence on L2 development is linked to learning conditions. Both of these types of research are discussed below. Because the target feature of the present study is the present third person singular, the following sections will focus on research that has explored the effects of aptitude on morphosyntax.

2.4.5 Aptitude as a Predictor of Ultimate Attainment

Studies that have explored the extent to which aptitude influences ultimate attainment when exposed to natural L2 conditions have attempted to shed light on the relationship between aptitude and age. Although there is agreement that aptitude affects ultimate attainment for older learners, previous research has demonstrated mixed findings as to whether aptitude affects the ultimate attainment of learners who started to acquire their L2 in early childhood (Abrahamsson & Hyltenstam, 2008; Bylund, Abrahamsson, & Hyltenstam, 2012; DeKeyser, 2000; DeKeyser, Alfi-Shabtay, & Ravid, 2010; Granena, 2014; Harley & Hart, 1997).

Harley and Hart (1997) was one of the first studies to examine the role of age in the relationship between aptitude and second language proficiency. The participants attended French immersion classrooms and differed as to whether they were exposed to the L2 from grade 1 (early immersion) or started in grade 7 (late immersion). The study showed that different aptitude components were associated with the L2 outcomes of the two groups. Specifically, for the early immersion students, a positive relationship was demonstrated between the learners' memory ability and their L2 performance, whereas for the late immersion group, the participants' language analytic ability was a significant predictor of their performance.

Other studies have explored the role of L2 aptitude under natural conditions while the participants lived in an L2-speaking country. For example, in DeKeyser (2000), the participants were Hungarian speakers in the USA. Their knowledge of morphosyntax was assessed by an auditory grammaticality judgement test (GJT), and their L2 aptitude was measured by a verbal analytical ability test administered in their L1. A relationship was found between aptitude and performance on the GJT for those participants who arrived in the USA after the age of 15. By contrast, there was no such relationship for those who arrived earlier. DeKeyser et al. (2010) obtained similar findings to DeKeyser (2000) with Russian speakers who used English as an L2 in the USA and Hebrew as an L2 in Israel. DeKeyser found that L2 aptitude influenced the ultimate attainment of adult L2 learners (18–40 years old), but not of learners younger than 18. Based on these findings, DeKeyser argued that because children employ implicit mechanisms in the L2 learning process, their aptitude (e.g. their language analytic ability) does not seem to affect their ultimate attainment. Conversely, as adult learners learn the L2 explicitly, their ultimate attainment is more likely to be influenced by their L2 aptitude.

Nonetheless, other studies, such as Abrahamsson and Hyltenstam (2008), Bylund et al. (2012) and Granena (2014), have demonstrated that aptitude is related to ultimate attainment for those participants exposed to the L2 from early childhood. In Abrahamsson and Hyltenstam (2008), the participants were L1 Spanish-L2 Swedish speakers who differed as to whether their L2 acquisition began at/before the age of 11 or at/after the age of 12. In Bylund et al. (2012), the participants' profiles were the same; however, they focused on learners whose L2 exposure started before the age of 12. Abrahamsson and Hyltenstam (2008) used auditory and written GJTs, and Bylund et al. (2012) administered a timed auditory GJT and an untimed cloze test. Both of these studies assessed phonetic memory, analytical ability, grammatical inferencing, sound recognition and sound-symbol correspondence using the Swansea Language Aptitude Tests (LAT) designed by Meara, Milton, and Lorenzo-Dus (2003). Contrary to DeKeyser's findings, they demonstrated that aptitude significantly predicted the ultimate attainment of early L2 learners.

Similarly, Granena (2014) examined whether aptitude affects ultimate morphosyntactic attainment for early L2 learners. The study employed a speeded-response and a non-speeded response GJT and the LLAMA aptitude test (Meara, 2005). The participants were Chinese speakers and L2 learners of Spanish living in Madrid. They had arrived in the L2 country early, when they were from 3 to 6 years old, or they had been born in Spain but had Chinese-speaking parents, and hence they had not been exposed to the L2 prior to the age of 3. The study revealed that aptitude was a significant predictor of learners' attainment, but only in constructions that required grammatical agreement and only on the non-speeded response GJT.

A possible reason for the mixed findings with regard to the role of L2 aptitude and ultimate attainment of early L2 learners might be related to what aptitude constructs different studies tapped into. For example, DeKeyser (2000) focused on verbal analytical ability, whereas the other three studies (Abrahamsson & Hyltenstam, 2008; Bylund et al., 2012;

Granena, 2014) used the LAT and the LLAMA aptitude test, which involve a variety of aptitude components. The contradictory findings about the relationship between aptitude and early starters' ultimate attainment might also be associated with the tests different studies utilized and the types of morphosyntactic features. Granena (2014) found that the type of outcome measure and the type of linguistic construction moderated the influence of aptitude on early starters' performance. Granena also explains that the type of assessment might be the reason why, in a previous study, Granena and Long (2013) found no relationship between aptitude and ultimate attainment on morphosyntactic elements. Granena argues that when aptitude is measured by a LLAMA test, that is mainly related to explicit cognitive processes, its relationship with morphosyntactic attainment might be better illuminated when GJTs are utilized rather than a combination of outcome measures, including an oral narrative, as in Granena and Long's (2013) study. All these methodological parameters (e.g. aptitude measures, type of assessment tool and type of linguistic construction) should be further explored before any conclusions are drawn about the impact of aptitude on early starters' ultimate attainment.

It is also worth emphasizing that those studies that explored aptitude as a predictor of ultimate attainment mainly focused on natural L2 learning conditions. The role of aptitude in foreign language classrooms is an under-researched area. A study that has delved into this issue is Saito (2017). In particular, Saito explored whether different aptitude constructs, measured by a LLAMA test (see section 2.4.4), were linked to different areas of students' second language speech achievement in terms of pronunciation, fluency, vocabulary and grammar. Saito's study was conducted in an EFL context involving Japanese speakers learning English as a second language with various proficiency levels. Saito found that rote and associative memory (LLAMA B) positively correlated with linguistic complexity and fluency; analytic ability (LLAMA F) was related to lexical richness; and phonemic coding (LLAMA E) was associated with pronunciation and the production of accurate morphology. Interestingly, no relationship was found between the sound recognition construct (LLAMA D) and any areas of learners' speech production. Saito attributed the lack of connection between sound recognition and oral production to the type of learning that LLAMA D may measure. In line with Granena (2013), Saito argues that the LLAMA D construct may be associated with more implicit learning mechanisms and is less likely to be employed in EFL conditions involving mainly explicit learning.

2.4.6 Aptitude as a Variable Interacting with Learning Conditions

Robinson's (2001a, 2002, 2007a, 2012) *Aptitude Complexes Hypothesis* provides underpinning for research on aptitude-treatment interaction. Robinson's model comprises four aptitude complexes that involve cognitive processes related to primary abilities such as *perceptual speed*, *pattern recognition*, *phonological working memory capacity, speed of processing in phonological working memory, working memory for text* and *speed of working memory for text, analogies of meaning* and *inferring word meaning,* *grammatical sensitivity* and *rote memory*. Robinson argues that these primary abilities entail *second-order abilities,* such as *noticing the gap*, *memory for contingent speech*, *memory for contingent text, deep semantic processing* and *metalinguistic rule rehearsal* (see Robinson, 2001a, 2002, 2007a, 2012). Robinson suggests that these abilities, or a combination of them, promote L2 development but their facilitative role is linked to specific learning conditions. For example, in Robinson's model, the extent to which learners notice and benefit from recasts is associated with (1) ability to notice the gap between an erroneous utterance and its reformulation and (2) memory for contingent speech (i.e. a reformulation has to be held in memory while comparing it with erroneous output). Robinson also expects that incidental learning from oral content involving a flood of linguistic features is related to memory for contingent speech and deep semantic processing, while incidental learning from written content is linked to memory for contingent text and deep semantic processing. Finally, explicit rule learning combined with written examples could be influenced by memory for contingent text and metalinguistic rule rehearsal.

In the second section of his framework, i.e. the *Ability Differentiation Hypothesis,* Robinson suggests that there is variation among L2 learners regarding their strengths and weaknesses in the cognitive ability involved in L2 aptitude. Due to these variations, learning conditions should be adapted to learners' ability in order to arrive at superior L2 outcomes. For example, recasting might be an effective intervention for learners with strengths in noticing the gap and memory for contingent speech, while for learners lacking these strengths, alternative focus-on-form techniques may be beneficial. Several empirical studies have sought to elucidate whether there is a relationship between individual differences in L2 aptitude and the effectiveness of different types of instruction (DeGraaff, 1997; Erlam, 2005; Hwu & Sun, 2012; Robinson, 1997; VanPatten, Collopy, Price, Borst, & Qualin, 2013).

Robinson (1997) explored the extent to which the potential benefits of four learning conditions are influenced by L2 aptitude measured by the MLAT (Carroll & Sapon, 1959). The learning conditions were classified into *instructed condition*, *rule-search condition*, *implicit condition* and *incidental condition*. In the *instructed condition*, first, the learners were provided with an explanation of rules, and second they were required to apply these rules to sentences (e.g. to answer questions regarding metalinguistic information about form). In the *rule-search condition*, the participants were instructed to look at sentences and find rules by themselves. In the *implicit condition*, the learners were presented with sentences and asked questions by the researcher (e.g. about the location of words), but without being given with any explicit or metalinguistic information. Finally, in the *incidental condition*, the learners read sentences and had to answer comprehension questions related to their content. The participants received feedback about the correctness of their answers in all of the conditions, apart from the rule-search condition where their answers could not be predicted. L2 learning was measured by a GJT and rule awareness was examined by a questionnaire asking the participants whether they had noticed or searched for rules and could verbalize them. Robinson found that all the experimental groups were influenced by aptitude, apart from the incidental learning group.

Similar to Robinson (1997), DeGraff (1997) also employed different instructional conditions in order to investigate their potential relationship with aptitude. Aptitude was operationalized as grammatical sensitivity and rote memory and it was measured by a Dutch version of the MLAT test and learners' ability to infer the meanings of novel words from the context. The L2 was an artificial language called eXperanto. In the explicit group, the learners' attention to linguistic elements was drawn by providing them with grammatical explanations and highlighting the target constructions. In contrast, no metalinguistic explanation was offered under the implicit condition. L2 learning was assessed by GJTs with and without time pressure and an untimed gap-filling task. DeGraff found significant correlations between L2 aptitude and both groups' performance on an immediate and a delayed posttest.

Erlam (2005) examined the relationship between individual differences in language analytic ability, phonetic coding ability and working memory and three different types of instruction, namely, *deductive instruction*, *inductive instruction* and *structured input instr*uction addressing direct object pronouns. *Deductive instruction* was coded as an explicit learning condition because it entailed rule explanation and learners' engagement with form-focused activities requiring the production of the target feature without time pressure. Hence, the participants had time to use the rules explained to them prior to the activities. These activities were followed by CF that aimed to draw their attention to the rules of the target construction. *Inductive instruction* was coded as an implicit learning condition. The participants performed activities and were expected to make their own hypotheses about the target feature without being presented with rules or any metalinguistic information. During *structured input instruction*, the learners were also presented with rules about the target feature, followed by input-based activities and consciousness-raising activities. The former required the participants to process both spoken and written input, including the target construction, whereas the latter asked them to identify errors. Neither of these activities involved production of the target feature. Erlam found that the explicit deductive instruction was less affected by aptitude, as opposed to (1) inductive instruction and (2) structured input instruction, both of whose benefits were moderated by the participants' L2 aptitude.

In a similar vein, Hwu and Sun (2012) examined the relationship between L2 aptitude and the effectiveness of two types of explicit instruction in facilitating the development of the Spanish verb gustar. The aptitude constructs they explored were analytical language ability and associative memory, measured by the MLAT and memory for text. The researchers employed two instructional conditions involving deduction and induction coded as explicit. Both groups were provided with an explanation of rules and metalinguistic information through instructional activities in the deductive condition, and through multiple-choice questions and metalinguistic feedback in the explicit inductive condition. L2 development was assessed by a pretest, an immediate posttest and a delayed posttest composed of tasks that required written sentence production and written sentence correction. The study indicated that the L2 gains of both groups were influenced by aptitude, and particularly by their memory for text.

Finally, VanPatten et al. (2013) conducted three experiments that differed with regard to the L2 in focus (i.e. Spanish, German and Russian for each experiment, respectively). The aim was to illuminate the relationship between the participants' grammatical sensitivity measured by the MLAT, the learners' processing time (i.e. length of time they needed in order to start processing the target feature accurately) and their improvement on the target construction. The linguistic area addressed referred to the flexibility of the target languages with respect to the order of words functioning as subjects or objects and whether they afforded both subject-verb-object (SVO) constructions and object-verb-subject (OVS) constructions. Based on the First-Noun Principle, the learners were expected to process the first noun as subject in both types of sentences. That would be the target-like option for the SVO constructions but not for the OVS ones. The participants were assigned into two groups, both receiving structured input of the target linguistic element; however, the two experimental conditions differed as to whether the structured input was combined with explicit information or not. The researchers found that the degree to which the students benefited from structured input positively correlated with their grammatical sensitivity only under the condition that involved explicit information. It should also be noted that this finding was demonstrated only in one of the three experiments, the one utilizing German as L2.

Overall, the studies presented above have shown that instructional conditions, regardless of whether they are coded by the researchers as explicit or implicit, are related to L2 aptitude. That is, learners with greater aptitude are likely to benefit more from both explicit and implicit instruction. However, these findings should be interpreted with caution. The factor that differentiates explicit from implicit learning is awareness; learning is implicit when learners are not aware of what they have learnt and explicit when awareness is implicated (DeKeyser, 2003; Rebuschat, 2013). Hence, an instructional condition devoid of the provision of rules and metalinguistic information does not necessarily entail implicit learning, as although the learners might not be able to verbalize a rule, they might have employed explicit cognitive mechanisms during the learning process and could be aware of their emerging L2 knowledge. Consequently, the instructional conditions operationalized as implicit in the studies above cannot exclude the possibility of explicit learning being involved, and they cannot provide robust evidence about the relationship of L2 aptitude with purely implicit vs explicit learning.

Another interesting parameter that should be taken into account is the impact of aptitude on the acquisition of linguistic constructions that differ in terms of salience, redundancy and/or rule complexity (Skehan, 2014a; Yalçın & Spada, 2016). For example, Yalçın and Spada (2016) examined the extent to which aptitude is associated with the acquisition of two linguistic areas that differed in rule complexity. The participants received four hours of instruction on the passive and the past progressive, with the former being coded as a more difficult structure than the latter. L2 gains were assessed by a pre-test and a post-test composed of written tasks, untimed GJTs and oral production tasks. L2 aptitude was measured by the LLAMA Aptitude test (Meara, 2005). Interestingly, the study revealed that different components of aptitude contributed to development of the two constructions. In particular, greater grammatical inferencing facilitated benefits related to the passive, whereas greater associative memory assisted improvement in the past progressive.

Within the framework of aptitude-treatment interaction, researchers have also explored the relationship of aptitude with L2 gains resulting from different types of CF (Li, 2013, 2015; Sheen, 2007; Trofimovich, Ammar, & Gatbonton, 2007; Yilmaz, 2013; Yilmaz & Koylu, 2016; Yilmaz, Granena, & Meyer, 2016; Yilmaz & Granena, 2015). These studies are discussed in greater detail in the following section.

2.4.7 L2 Aptitude and Corrective Feedback

Although a plethora of research has demonstrated the benefits of CF, the extent to which the effectiveness of different CF techniques is associated with aptitude constructs such as language analytic ability constitutes an interesting area of research. Although a relationship has been found between explicit feedback and learners' aptitude (Yilmaz & Koylu, 2016;

Yilmaz, Granena, & Meyer, 2016; Yilmaz & Granena, 2015), research on recasts and aptitude has demonstrated contradictory findings. In particular, Sheen (2007) and Yilmaz and Granena (2015) found no relationship between the efficacy of recasts and aptitude, while Li (2013) and Trofimovich et al. (2007) showed a positive relationship.

Sheen (2007) compared two types of feedback (i.e. recasts vs metalinguistic correction) in terms of their effectiveness in promoting the acquisition of English articles. She also examined the extent to which learners' language analytic ability moderated the effect of the two types of feedback in focus. Implicit feedback provided in the form of recasts was operationalized as (1) reformulations of learners' full utterances or (2) reformulations of only the erroneous parts of their utterances. In contrast, metalinguistic correction aimed to draw learners' attention to their errors more explicitly by providing both the target-like linguistic construction and metalinguistic information. During the two treatment tasks, the participants had to read and narrate a story; the experimental groups received recasts or metalinguistic feedback in response to errors related to articles, while the control group were given no correction. L2 development was measured by a pre-test, a post-test and a delayed post-test composed of three tasks. The first task involved speeded dictation, during which the participants had to produce articles under time pressure without the possibility to revise their output. The second assessment task was a writing test during which the learners had to write a story based on pictures accompanied by prompts.

Finally, the students were administered an error correction test that asked them to correct sentences with the target linguistic construction.

Compared to the control group, Sheen's study found that metalinguistic feedback was significantly more effective than recasts on both the immediate and the delayed post-test. The recast and control groups did not differ significantly in their performance on any of the tests. The study also found a significant positive correlation only between the students' analytic ability and their improvement after receiving metalinguistic feedback on both the immediate and the delayed post-test. Meanwhile there was no relationship between language analytic ability and recasts, i.e. even students with strong analytic ability did not benefit from recasts.

Similar to Sheen (2007), Yilmaz and Granena (2015) explored the relationship between L2 aptitude with explicit and implicit feedback. In their study, the participants performed a guided oral production task, a spot-the-difference task and a story retelling task, and they received either explicit correction or recasts on the English indefinite article. L2 development was assessed by a pre-test, an immediate post-test and a delayed post-test, using the same tasks as those employed during the treatment. L2 aptitude was measured by three of the components of the LLAMA Aptitude Test (Meara, 2005): LLAMA B, LLAMA E and LLAMA F, referred to as explicit aptitude (see section 2.4.4). In line with Sheen (2007), Yilmaz and Granena (2015) demonstrated that the learners' improvement on the immediate post-test was associated with L2 aptitude, but only under the explicit condition. No relationship was found between recasts and any of the components of aptitude on either the immediate or the delayed posttest.

Overall, Sheen (2007) and Yilmaz and Granena (2015) found (1) a positive relationship between the participants' L2 aptitude and improvement after receiving explicit feedback and (2) no relationship between aptitude and implicit feedback operationalized as recasts. Sheen (2007) argues that a possible reason why no positive correlation was found between recasts and learners' language analytic ability in her study could be a lack of students' awareness of English articles. Sheen points out that language analytic ability might facilitate learning when students receive explicit feedback offering them metalinguistic information about a target feature, rather than when they are provided with a target-like model as in recasts.

However, unlike Sheen (2007) and Yilmaz and Granena (2015), Trofimovich et al. (2007) and Li (2013) obtained a different finding with respect to the relationship between recasts and aptitude. Trofimovich et al. (2007) examined whether the extent to which students noticed and learned from recasts was associated with their analytical ability and attention control, i.e. the ability to allocate attention among various aspects of language. The participants were asked to perform a picture description task during which they received morphosyntactic recasts, lexical recasts or both. After receiving a recast, noticing was measured by asking the learners whether they noticed any differences between their own descriptions and the descriptions provided to them afterwards. L2 development was gauged by a pre-test, a post-test and a delayed post-test also requiring picture descriptions. The researchers found that recasts benefited the learners, as they increased their scores from the pre-test to the immediate and delayed post-tests; however, the benefits of morphosyntactic recasts on the delayed post-test were mediated by individual differences in analytical ability. In other words, only learners with greater analytical ability had long-term gains from recasts. The study also found that attention control was associated with the effectiveness of morphosyntactic and lexical recasts.

Li (2013) obtained a similar finding when he explored whether the efficacy of recasts and metalinguistic correction was linked to language analytic ability. The participants engaged in dyadic NS-NNS interactions and were assigned to one of three experimental conditions. The recast group delivered recasts that were partial, didactic and ended in a falling tone, the metalinguistic group were provided with a target-like construction accompanied by a metalinguistic clue, while the control group received no feedback. The recasts and metalinguistic feedback addressed Chinese classifiers, which refer to words used between determiners (e.g. numbers, demonstratives, quantifiers). L2 gains were measured by a pretest-posttest composed of a GJT and an elicited imitation test. The participants' language analytic ability was assessed by the Words in Sentences subtest of the MLAT (Carroll & Sapon, 2002). In Li's study, the highest scores on the posttests were achieved by the metalinguistic group, followed by the recast group, which in turn outperformed the control group. As for language analytic ability, it affected L2 benefits on a delayed grammaticality judgement posttest only under the recast condition.

A possible explanation for the mixed findings regarding the relationship between recasts and explicit components of L2 aptitude (e.g. language analytic ability) is that it is influenced by the type of linguistic construction. This is a presumption reinforced by Yilmaz's (2013) findings. Yilmaz investigated (1) the effectiveness of explicit correction and recasts and (2) whether their efficacy was associated with language analytic ability. Recasts were operationalized as "the target-like reformulation of the erroneous segment of the learner's production" (p. 352), while explicit correction involved direct rejection of learners' utterances accompanied only by the provision of the target-like construction without metalinguistic information. The participants were native speakers of English who had not previously been exposed to Turkish. The learners engaged in a one-way information gap task and, depending on their group, they received recasts, explicit correction or no feedback on the two target constructions, which were the Turkish plural morpheme and the Turkish locative case morpheme. As the learners had no prior knowledge of Turkish, a pre-test was not administered. After the treatment tasks, the students completed an immediate post-test consisting of an oral production, comprehension and recognition test, and two weeks later, they were administered a delayed post-test with the same components. Their language analytic ability was measured by the LLAMA Language Aptitude Tests (Meara, 2005). Interestingly, Yilmaz demonstrated that the relationship between language analytic ability and recasts was moderated by the target construction. In particular, when recasts addressed the plural, only learners with high language analytic ability benefited from recasts, whereas those with low language analytic ability did not. In contrast, when the target construction was the Turkish locative, the performance of the recast group was the same as the control group, regardless of the students' language analytic ability.

Therefore, before drawing any conclusions about whether the effectiveness of recasts is related to learners' L2 aptitude (e.g. language analytic ability), the type of target linguistic construction should be examined. Li (2013) and Yilmaz (2013) argue that when recasts target more salient and simple constructions, learners equipped with higher language analytic ability might employ their internal resources and, as a result, benefit from more implicit feedback such as recasts. Thus, language analytic ability positively influenced recasts addressing Chinese classifiers that had more obvious and less complex form-meaning mappings (Li, 2013), and the Turkish plural, a relatively simple construction with only two allomorphs (Yilmaz, 2013). Meanwhile learners with higher aptitude did not benefit from recasts when the target feature was non-salient, relatively complex English articles in Sheen (2007) and in Yilmaz and Granena (2015), and Turkish locatives in Yilmaz (2013). Yilmaz explains that, unlike the plural, the Turkish locative is a less salient and a more complex construction composed of four allomorphs. According to Li (2013), recasts targeting less salient constructions related to complex form-meaning mappings might not encourage learners to exploit their language analytic ability and extract relevant rules on their own. But learners might resort to their language analytic ability when they have to process recasts that address salient features with clear and simple form-meaning connections.

Previous research has also shown that learners' level of proficiency is another factor that impacts on how aptitude correlates with recasts. In particular, it has been found that aptitude constructs are differentially related to the efficacy of recasts for low level and high level students. For example, in Li (2015), learners of Chinese were divided into beginners and advanced-level students, and they received recasts on Chinese classifiers. L2 gains were captured by a grammaticality judgement and an elicited imitation pretest, immediate posttest and delayed posttest. The participants' working memory was measured by a listening span test (Waters & Caplan, 1996) and language analytic ability was assessed by the words in sentences subtest of the MLAT. Li found that both groups exhibited improved performance that was sustained; however, each group took advantage of a different aptitude construct: the beginners employed their language analytic ability, while the advanced learners benefited from their working memory.

2.4.8 Summary

In summary, the studies presented in the previous sections investigated the degree to which aptitude is linked to ultimate attainment and the effectiveness of different instructional conditions, including explicit and implicit CF. L2 aptitude appears to be associated (1) with the ultimate attainment of both older learners and those who started in early childhood, (2) with learners' second language achievements in an EFL setting, (3) with the benefits of different types of instruction (e.g. deductive and inductive) and (4) with the efficacy of different types of CF (e.g. explicit and implicit). Research has also indicated that the relationship between aptitude and morphosyntactic development might be associated with the type of linguistic construction, both in studies of ultimate attainment (see Granena, 2014) and in studies of instructional conditions and CF (see Yalçın & Spada, 2016; Yilmaz, 2013). Learners’ level of proficiency has also been found to play an important role when examining the effects of aptitude (Li, 2015).

Although SLA researchers have mainly focused on the relationship between CF and constructs such as WMC and language analytic ability, other aptitude components have not been sufficiently explored in SLA research.[3] Interestingly, phonetic coding ability has been found to influence the effectiveness of explicit feedback (Yilmaz & Koylu, 2016); however, its relationship with recasts has not been illuminated. Moreover, the role of variables such as task complexity (see section 2.3) and mode of interaction (see section 2.2) in the relationship between aptitude and the efficacy of CF has not been examined, especially for young learners who have not yet had ample exposure to the L2 (i.e., learners of English in an EFL context).

---

[3] Yilmaz and Granena (2015) examined the effects of a combination of aptitude constructs, referred to as explicit aptitude (LLAMA B, E, and F), without showing the role of these components separately.

To address these gaps, the current thesis focused on students in the first stages of L2 development whose ages ranged from 10.5 to 13 years. The participants received recasts on the present third person singular, a non-salient feature, while performing cognitively simple and complex tasks (Study 1, chapter 3), and while interacting in the FTF and SCMC modes (Study 2, chapter 4). Moreover, the relationship of recasts with aptitude constructs such as ability to associate sounds with their symbols, ability to recognize oral patterns and language analytic ability was addressed in both studies. Now we first turn to a description of Study 1, where the main focus was recasts, task complexity and aptitude, followed by a report on Study 2, which explored the links among recasts, mode of interaction and aptitude.

CHAPTER 3: STUDY 1

3.1. Aims and rationale

The aim of the first study was to explore the combined effects of task complexity and recasts on modified output and development in the knowledge of the present third person singular, and whether these relationships are influenced by L2 aptitude. It also sought to investigate possible correlations of modified output with L2 outcomes. As explained earlier, in chapter 2, previous literature has investigated the efficacy of corrective feedback (CF) supplied during cognitively simple and complex tasks (see section 2.3.3); however, the following gaps have been identified.

First, the relationship of modified output with subsequent learning is still unclear due to the mixed findings of previous studies. It is possible that the extent to which modified output is related to L2 development may be influenced by learning conditions. Apart from Révész, Sachs, and Mackey (2011), no other study has attempted to illuminate this issue.

Second, output modification and subsequent development in knowledge of the present third person singular, a non-salient and communicatively redundant linguistic feature, has not received sufficient attention from researchers working on task effects. It could be argued that when a linguistic construction is not needed for task completion and meaning can be conveyed successfully without its use, increasing the cognitive demands of a task may lead to less attention being allocated to the linguistic target, especially when it is of low salience and the students are young learners in the first developmental stages.

Third, it should be emphasized that young learners, the target population in the present study, are an under-researched population in SLA. Given that the present third person singular is a feature prone to fossilisation in adulthood (Han, 2013), it is important to identify learning conditions that promote interlanguage development in young learners' knowledge of this feature.

Another research gap relates to the type of recasts employed. Most studies utilizing recasts have usually operationalized them as implicit feedback (e.g. full recasts), although some studies have demonstrated that explicit recasts are more effective than implicit ones in drawing learners' attention to linguistic elements (see sections 2.1.4 and 2.1.5). The current study used explicit, partial recasts in order to increase the salience of the present third person singular and possibly facilitate learning for students with very limited background knowledge of the linguistic target.

Finally, the present study intended to reveal potential associations between L2 aptitude and the effects of recasts on L2 outcomes, and whether these relationships are affected by task complexity. Previous research has exhibited that aptitude is related to the efficacy of recasts (see section 2.4.7); however, it has not been explored whether this relationship transfers to both simple and complex tasks. Although Robinson (2011) claims that individual differences in cognitive abilities "will increasingly differentiate learning and performance as tasks increase in complexity" (p. 19), the role of such abilities has not yet been elucidated for complex tasks.

3.2 Research Questions and Hypotheses

To fill these gaps, the current study focused on a non-salient feature (i.e. present third person singular); it was conducted in an EFL context with young learners in the first stages of SLA, and it employed explicit rather than implicit recasts. Also, it did not only explore the extent to which recasts and task complexity facilitated modified output and L2 gains, it also investigated how these relationships were influenced by learners' aptitude. Hence, three sets of research questions were formulated for the first study: the first set was concerned with links among task complexity, modified output and L2 development. The second set investigated the relationships of aptitude to modified output and L2 outcomes, regardless of task complexity. Finally, the third set addressed how task complexity affects the relationship of aptitude to modified output and L2 gains.

Task complexity, modified output and L2 development

1. To what extent does task complexity affect the amount of modified output after recasts targeting the present third person singular?
2. What are the combined effects of task complexity and recasts on promoting knowledge of the present third person singular?
3. To what extent does task complexity influence the relationship between modified output and development in knowledge of the present third person singular, while recasts remain constant?

Aptitude, modified output, and L2 development

4. To what extent does aptitude relate to the amount of modified output after recasts targeting the present third person singular?
5. To what extent does aptitude relate to the effectiveness of recasts in promoting knowledge of the present third person singular?

Task complexity, aptitude, modified output and L2 development

6. To what extent does task complexity influence the relationship between L2 aptitude and modified output after recasts targeting the present third person singular?
7. To what extent does task complexity influence the relationship between L2 aptitude and development in knowledge of the present third person singular, when recasts are held constant?

Task complexity was operationalized as +/- reasoning demands (see section 3.4.4). Drawing on Skehan's framework, it was presumed that the increased cognitive demands posed by +reasoning condition would hinder deeper processing of the linguistic element addressed by the recasts, considering that (1) the linguistic target was a non-salient, redundant feature not easily noticed by L2 learners and (2) the participants were young learners in the first stages of development, with very limited prior knowledge of the target feature. That is to say, recasts delivered while performing tasks with greater reasoning demands would probably not enable the creation of initial mental representations of the linguistic target as the students would probably devote greater attention to the content of their utterances. In contrast, the simple condition that involved the provision of static information without imposing reasoning demands was expected to free up attentional resources; and as a result, the participants would probably be able to allocate greater attention to linguistic encoding and to CF in the form of recasts addressing a non-salient, redundant linguistic construction that had not yet been internalized.

Hence, based on the Limited Capacity Model, the hypotheses of the current study were that simple tasks would lead to (1) more modified output after receiving recasts and (2) greater development in knowledge of the present third person singular, given the attentional limitations of learners, the low salience and redundancy of the target construction, and the developmental stage of the students regarding the target feature. As for the third research question, no hypothesis was formulated as previous studies had yielded mixed findings as to whether modified output was a predictor of subsequent L2 development (see section 2.1.4). Although Révész, Sachs, and Mackey (2011) found that task complexity influenced whether modified output predicted L2 learning, their task manipulation was different from this study, and hence no directional hypothesis can be formulated drawing on this research.

Regarding aptitude (research questions 4–7), it was expected that it would positively correlate with modified output and L2 development, especially under the complex condition. In other words, learners with higher aptitude were hypothesised to benefit more from recasts, especially when they engaged in tasks imposing greater cognitive demands. This speculation is in line with Robinson's (2011) prognosis that cognitive ability will lead to greater variation in L2 outcomes when learners carry out more cognitively demanding tasks. Two aptitude constructs that were expected to positively correlate with participants' improvement were the ability to recognize oral patterns and the ability to associate sounds with symbols, as the learners had to process oral recasts targeting the various allomorphs of the present third person singular (i.e. /s/, /z/ and /əz/). Another aptitude construct that was expected to be linked to the efficacy of recasts was the participants' grammatical inference ability, as the target construction was a grammatical feature and the feedback did not involve metalinguistic explanations. That is, learners had to infer the target pattern using their own cognitive resources. Hence, the most relevant aptitude constructs appeared to be the ones measured by the LLAMA D, E and F tests.

3.3 Pilot study

The purpose of the pilot study was to gain preliminary insights into the combined effects of recasts and task type on development in knowledge of the present third person singular, and to indicate potential weaknesses of the treatment and assessment tasks. Task type was operationalized as information transmission versus decision-making tasks. The current section discusses the design and the materials used for the pilot study, and it explains what changes were made in study 1, based on the findings of the pilot study.

3.3.1 Design

The pilot study employed a pretest-immediate posttest design and three treatment sessions (see Table 4). The participants were randomly assigned to one of two experimental groups: one group of students engaged in information transmission tasks and the other in decision-making tasks. Recasts addressing the present third person singular were delivered in both conditions. The combined effects of recasts and task type were explored by measuring the learners’ development from pretest to posttest.

TABLE 4 DESIGN OF THE PILOT STUDY

	Information transmission N=10	Decision-making N=10
Pretest	Oral production	
	Elicited imitation	
	Written production	
Treatment 1	Task 1	
Treatment 2	Task 2	
Treatment 3	Task 3	
Posttest	Oral production	
	Elicited imitation	
	Written production	

3.3.2 Participants

The participants of the pilot study were 20 learners of English from a summer language school in the UK. The school employed pedagogical materials following the principles of task-based language teaching (TBLT). The background information of the participants is given in Table 5 Overall, there were 7 males and 13 females, and their ages varied from 14 to 22 years (M=16.75, SD=.2.57). Based on their placement in the summer school, their proficiency level was pre-intermediate/intermediate. Their L1s were Russian, French, Portuguese and Chinese. The participants were in an ESL context during the experiment; however, prior to the summer classes, they attended English in an EFL context (i.e. in a school in a non-English speaking country). The length of their stay in the UK varied from two to four weeks.

TABLE 5 PARTICIPANTS OF STUDY 1

Group	N	Gender	Age		
			M (SD)	Median	IQR
Information Transmission	10	4 males, 6 females	16.80 (2.52)	16.00	4.50
Decision-making	10	3 males, 7 females	16.70 (2.75)	15.50	4.50

3.3.3 Treatment

The treatment involved three sessions comprising three oral decision-making tasks and three oral information transmission tasks for the experimental groups, respectively. The participants performed the tasks individually and they received recasts targeting the present third person singular. The decision-making tasks required the learners to make decisions and justify them by using information provided by the task, while the information transmission tasks asked the participants to give information related to a real-life context. Specifically, in the first decision-making task, the learners were the owners of a company that made a profit and they decided to buy their employees gifts in order to reward them for their hard work. The participants/ owners were given a list of potential gifts and tables with information about the weekend schedules of their employees. Based on this information, the participants had to choose the best gift for each employee and justify their decisions. For example, a potential justification for choosing trainers as a gift could be that an employee usually goes jogging at the weekend. The first task of the information transmission group involved the same context as the first decision-making task; however, the role of the students was different. In particular, during information transmission tasks, the participants were secretaries in a company and had to give information to the owner of that company about the employees' schedules in order to help her choose the best gifts for them and reward them for their hard work. Hence, the students were only given tables with information, but no list of gifts.

A similar manipulation was realized in the second treatment task of each group. Specifically, in the second task of the decision-making condition, the learners were employment counsellors and had to use information related to their clients' schedules in order to suggest the most suitable jobs to them. For example, a possible job for one of the clients could be as a fashion designer because she draws at the weekend. In the second task of the information transmission group, the learners worked as secretaries in an employment agency and had to give information to the owner of the agency about what the clients usually did at the weekend so that their employer could make decisions about the best job for each client.

In the third task of the decision-making group, the learners were the owners of a travel agency and, similar to the other two tasks, they had to use information about their clients' schedules in order to choose ideal travel destinations for them. For example, a good travel destination for one client could be Ibiza because he goes to parties at the weekend. In the third task of the information transmission group, the learners worked as secretaries in a tourist agency and had to relay information about what the clients usually did at the weekend so that their employer could make decisions about the best travel destinations. Overall, the context of each decision-making task was the same as the context of the information transmission tasks, but the role of the participants was different (i.e. owner of a company/ employer making a decision or a secretary giving information). The tables with habits were the same for both conditions so as to elicit the same verbs from both groups.

The pilot study indicated that the three treatment tasks were suitable for eliciting the target construction in both the decision-making and information transmission conditions. Nevertheless, these tasks were not employed in study 1 as they appeared to encourage the learners in the decision-making group to produce a very limited number of verbs. In particular, most of the participants in the decision-making condition justified why they chose a specific gift, job or destination by expressing the characters' preferences, and they mainly used verbs such as like, love and enjoy (e.g. 'I'll buy a cookbook for Kate because she like cook in the morning'). Meanwhile, the information transmission group produced a wide range of verbs by describing the actions depicted in the visual images (e.g. 'Kate usually cook in the morning'). Therefore, the two conditions were not comparable as the decision-making group received recasts on a limited number of verbs, whereas the information transmission group were provided with recasts targeting a wide variety of verbs. This discrepancy could confound the learners' potential development from pretest to posttest and, consequently, impede any conclusions to be drawn about the combined effects of recasts and task type. The researcher decided, therefore, that modifications to the treatment tasks of both conditions should entail the production of a wide variety of verbs in order to elicit recasts that would provide *balanced* rather than *skewed input*. Based on McDonough and Nekrasova-Becker's (2014) findings, it was assumed that skewed input containing exemplars from a limited number of verbs (e.g. like and enjoy) would not enable learners to generalize the pattern of the present third person singular in novel verbs. On the contrary, balanced input involving exemplars from a variety of verbs would lead to such generalizations and thereby facilitate interlanguage development. Hence, different treatment tasks were employed in study 1 to encourage the learners in both groups to focus on habits rather than preferences, and to use a wide range of verbs (see section 3.4.4). It should also be noted that post-task questionnaires were not used in the pilot study as potential differences between the two treatment tasks in terms of their complexity were not considered at that stage. In the first study, the decision-making tasks were judged to be cognitively more demanding than the information transmission tasks and, consequently, task complexity rather than task type was used as a factor differentiating the two treatment groups. A post-task questionnaire was also utilized to validate the construct of task complexity (see section 3.4.5).

3.3.4 Outcome measures

The outcome measures of the pilot study were administered individually and comprised an oral production test, an elicited imitation (EI) test and a written production test. The oral and written production tests involved visual images showing habits, and their aim was to elicit the present third person singular in the oral and written modes, respectively. The pilot study showed that both assessment tasks prompted the production of the present third person singular and the use of a variety of verbs and allomorphs. However, a few pictures were not clear to the students or they encouraged the production of the third person plural rather than the third person singular, as they depicted two characters. These pictures were replaced in studies 1 and 2 (see section 3.4.7).

For the EI test, the learners listened to both grammatical and ungrammatical sentences involving the target feature or distractors. First, they had to choose a picture relevant to the sentence in order to process its content; and second, they were asked to produce the sentence in correct English within five seconds. The main focus was on the production of the present third person singular while other errors were ignored in the analyses. The reliability of the EI test designed for the pilot study was examined by calculating Cronbach's Alpha and this was found to be low (no more than α = .500). The changes made in study 1 were related to the number of syllables, grammatical and ungrammatical verbs, choice of pictures, and the time the learners had to choose a picture and produce a sentence. In particular, in the EI test employed in the pilot study, the number of syllables was not the same for all the sentences, possibly resulting in a greater cognitive burden being imposed by sentences with more syllables. Second, the grammatical and ungrammatical verbs were different, hence any discrepancies between them could be attributed to item factors rather than the grammaticality or ungrammaticality of utterances. Third, some pictures were not clear to the students. Finally, some participants needed more than five seconds to produce sentences. Taking these factors into consideration, a novel EI test was designed for the first study in order to circumvent the weaknesses of the one utilized in the pilot study (see section 3.4.7).

After the changes described above, both the treatment and assessment tasks were also piloted on native speakers of English (N=4) and Greek young learners (N=7) whose ages ranged from 10.5 to 13 years. The participants of the pilot study presented above had very different characteristics from the young Greek learners who took part in the final study (e.g. they were older, they had different L1s, their proficiency levels were higher). Hence, piloting the research materials in a population similar to the one participating in the first study of the current thesis was considered crucial. Piloting the tasks and tests on the Greek learners indicated that no further changes were needed, and hence the materials were judged suitable for study 1.

3.4 Research Design and Methodology of Study 1

The aims and research questions of the first study are presented in sections 3.1 and 3.2. The current section is devoted to the design and methodology, followed by the results, for each research question and a detailed discussion.

3.4.1 Design

Study 1 employed a pretest-posttest design. Based on an initial proficiency test, 60 participants were selected to take part in the study and they were randomly assigned to one of two experimental groups (see Table 6). During the treatment, both groups engaged in face-to-face interaction with the researcher; however, they differed as to whether they performed information transmission tasks (i.e. no reasoning demands, simple version) or decision-making tasks (i.e. +reasoning demands, complex version). In both conditions (information transmission and decision-making), the learners carried out two treatment tasks and received interrogative recasts from the researcher in response to errors related to the present third person singular. Each treatment task was followed by a post-task questionnaire. The aim of the questionnaire was to gain information about the learners' perceived level of mental effort required by the tasks, and thereby validate the construct of task complexity. Development in knowledge of the target feature was assessed by three outcome measures: an oral production test, an elicited imitation (EI) test and a written production test. L2 aptitude was measured by the LLAMA test (Meara, 2005).

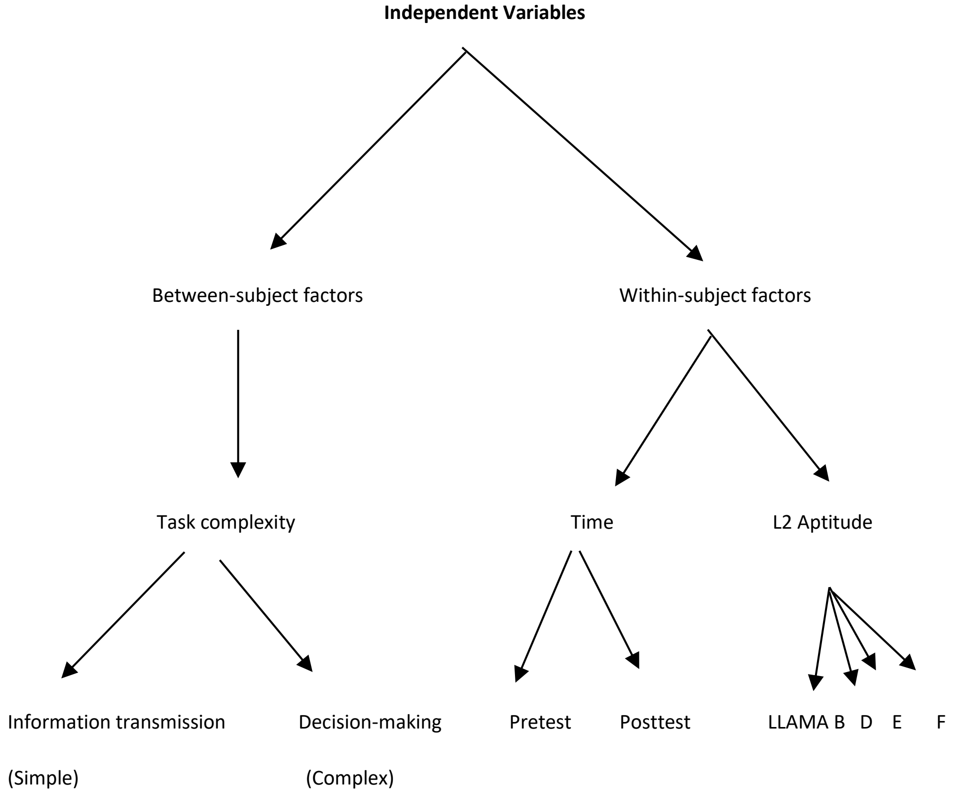
TABLE 6 DESIGN OF STUDY 1

	Simple task condition (Information transmission) N=30	Complex task condition (Decision-making) N=30
Proficiency test	ISE Foundation (A2) from Trinity College London	
Pretest	Oral production	
	Elicited imitation	
	Written production	
Treatment 1	Task 1	
Treatment 2	Task 2	
Posttest	Oral production	
	Elicited imitation	
	Written production	
L2 Aptitude	LLAMA B, D, E, F	

The first set of research questions was related to task complexity, modified output and L2 development. The first research question was addressed by comparing the amount of modified output produced by the two groups (i.e. information transmission and decision-making groups). The second research question was explored by comparing the gain scores of the two groups on each outcome measure. For the third research question, the relationship between L2 development and modified output was examined for the complex and simple conditions. The second set of questions referred to aptitude, modified output and L2 development. In this set, the fourth and fifth research questions were answered by investigating potential correlations between (1) aptitude and modified output and (2) aptitude and gain scores achieved by both groups regardless of task complexity. Finally, the third set of research questions involved task complexity, aptitude, modified output and L2 development. In particular, the sixth and the seventh research questions were explored by calculating correlational analyses between (1) L2 aptitude and modified output and (2) L2 aptitude and development. These correlations were run for each group separately in order to elucidate whether they were influenced by task complexity.

In sum, the study had (1) one between-subjects factor, task complexity, which involved two levels: (2) time as a within-subjects factor, which in turn consisted of two levels – a pretest and a posttest, and (3) two dependent variables – modified output and L2 development gauged by three tests. The study also included a (4) within-subjects factor, aptitude, which had four levels – LLAMA B, LLAMA, D, LLAMA E and LLAMA F (Meara, 2005). Figure 1 shows a schematic representation of the research design.

FIGURE 1 RESEARCH DESIGN OF STUDY 1



Independent Variables:

-Task complexity

-Learners' L2 aptitude

-Time of the test (pre- and post-test)

Dependent Variables:

-Amount of modified output

-L2 development from pre- to post-test comprising oral production, elicited imitation and written production.

3.4.2 Participants

The initial pool of participants was 160 L2 learners enrolled in various language schools in Greece. The students received instruction based on the communicative approach adopted by both state and private language schools in Greece. Of these participants, 60 elementary or pre-intermediate EFL learners were considered eligible to participate in the study. The selection criteria were determined based on the performance of the students (1) on a proficiency test – ISE Foundation (A2) from Trinity College London and (2) on pretests. The proficiency test was employed in order to obtain evidence that a student's level was elementary or pre-intermediate. Students' prior knowledge of the target feature was assessed by the pretests, and those who scored more than 35 per cent on one of the outcome measures were excluded from the study to avoid ceiling effects. The next highest score after 35 per cent was 50 per cent.

Background information about the final pool of participants, both groups, is presented in Table 7. Overall, there were 26 female and 34 male learners, and their ages ranged from 10.5 to 13 years (M=11.46, SD=.82). They were all native speakers of Greek, with the exception of two students who were bilingual speakers of Greek, one Albanian and one Romanian born in Greece. The length of learning English prior to the study varied from 2 to 8.5 years (M=4.65, SD=1.17). The majority of the participants also reported that they were learning French and German as a second language (N=48). None of the participants had lived in an English-speaking country before the study. A series of Mann-Whitney U tests run on the factors of age, length of learning English prior to the study and performance on the proficiency test confirmed that the two groups were comparable with respect to these variables (Mann-Whitney U = 402.50, p = .442 for age; Mann-Whitney U = 430.50, p = .764 for length of learning English before the research project; Mann-Whitney U = 445.50, p = .946 for their performance on a listening test). The effect sizes were r = .09, r = .03 and r = .00, respectively (section 3.4.12 explains the formula used to calculate the effect sizes). Therefore, the two groups were comparable in terms of their ages, numbers of years they had studied English and levels of proficiency.

TABLE 7 DESCRIPTIVE STATISTICS FOR BACKGROUND INFORMATION BY GROUP

Group	*N*	*Gender*	*Age in years*			*Length of previous English study in years*			*Proficiency*		
			M(SD)	Median	IQR	M(SD)	Median	IQR	M(SD)	Median	IQR
Simple Tasks	30	18 M 12 F	11.55 (.93)	11.00	1.00	4.70 (1.27)	4.25	2.00	5.83 (2.13)	5.00	3.00
Complex Tasks	30	16 M 14 F	11.38 (.71)	11.00	1.00	4.60 (1.09)	4.25	1.00	5.80 (2.05)	5.00	3.00

Note: IQR=Interquartile range, M=male, F=female

3.4.3 The Target Linguistic Construction

The target linguistic feature was the present third person singular, a construction denoting grammatical tense and agreement. According to Swan (2005), the present simple is typically used in the following contexts:

1. “To talk about permanent situations, or about things that happen regularly, repeatedly or all the time” (p. 450).
   E.g. Jack plays football on Wednesdays.
2. For things occurring “around the present” when non-progressive verbs are used such as like, believe (p. 450).
   E.g. I like your T-shirt.
3. To talk about the future.
   a. For timetables (e.g. The bus leaves at 8.00 a.m.)
   b. In subordinate clauses (e.g. I’ll explain this to you when I call you.).
   c. In suggestions (e.g. Why don’t we go to the theatre?).
4. To “talk about series of completed actions and events” (p. 451).
   E.g. So I call Jane and she asks me whether this is possible.
   Swan explains that this category involves the use of the present simple for stories, commentaries and instructions or demonstrations.

In the current study, the treatment tasks and tests were designed to elicit prototypical usage of the present simple, i.e. “to talk about permanent situations, or about things that happen regularly, repeatedly or all the time” (Swan, p. 450) (see sections 3.4.4 and 3.4.7).

One source of difficulty in acquiring the present simple is that it can be used in several contexts conveying different meanings (e.g. habits, future timetables, completed events of commentaries). Hence, deciding whether to use the present simple or other constructions such as the present progressive or future tenses poses a great challenge for L2 learners. With regard to students who are Greek speakers and L2 learners of English studying the language in an EFL context, acquisition of the present simple is challenging as a result of L1 interference. In particular, due to the lack of distinction between the present simple and progressive in Greek, students whose L1 is Greek may use the two tenses interchangeably, and often erroneously. The example below (3.1) demonstrates that the progressive aspect of the present tense does not exist in Greek. Specifically, the verb remains the same in the Greek sentences a and b.

(3.1)
a. Η Ελένη ***παίζει*** πιάνο τα Σάββατα.
"Helen *plays* the piano on Saturdays".
b. Η Ελένη ***παίζει*** πιάνο τώρα.
"Helen *is playing* the piano now".

Another difficulty with the present third person singular is that it involves a bound morpheme realized via three phonological alternations – the consonantal, non-syllabic allomorphs /s/, /z/, and a syllabic allomorph that contains a schwa vowel /əz/. Although these sounds can be produced by L1 Greek speakers, they are not easily noticed by L2 learners for two main reasons. First, as none of these allomorphs are stressed, they are classified as physically *non-salient* (i.e. not easily heard during oral interaction) (Goldschneider & DeKeyser, 2001; VanPatten, 1996); and consequently, they remain unattended. Second, their communicative value is low. VanPatten (1996) explains that communicative value refers to the contribution of a linguistic construction to the meaning of an utterance, and it is related to (1) its *inherent semantic value* and (2) *redundancy*. Regarding allomorphs of the present third person singular, although they have inherent semantic value, i.e. the semantic notion of the present third person singular, they are communicatively redundant (VanPatten, 1996). That is to say, the meaning of the present third person singular can be conveyed successfully from subject noun phrases, and hence the use of verbal inflection is redundant.[4] Due to its redundancy, it is more likely that the present third person singular remains unprocessed in oral interaction than other linguistic items that are necessary for successful communication. VanPatten argues that learners are more likely to process morphological elements with high communicative value in comparison to those that do not contribute to the meaning of an utterance.

Regarding CF techniques, they may not always be successful in drawing learners' attention to non-salient or redundant linguistic elements. For example, while engaging in meaningful interaction, L2 learners may not attend to CF addressing errors in verb-final allomorphs /s/, /z/, and /əz/ employed to mark the present third person singular, especially when more implicit feedback techniques are used such as recasts (see Yang & Lyster (2010) and Yilmaz (2012) with regard to CF and non-salient constructions). Long (2007) has also argued that unobtrusive recasts may not be effective in drawing learners' attention to such linguistic features.

3.4.4 Treatment Tasks

As mentioned earlier, the participants were randomly assigned to two experimental groups: The first group performed *decision-making tasks*,

[4] Subject noun phrases are obligatory in English apart from certain utterances related to colloquial English (e.g. Feels good).

whereas the second group carried out *information transmission tasks.* In both conditions, the tasks involved the habits of characters in order to elicit prototypical usage of the target construction, which was the present third person singular (see section 3.4.3). In particular, in the first decision-making task, each participant was the administrator of a building and the researcher was his/her assistant. They had found several items left by a transportation company at the entrance to their residence and the owners of these items were the tenants of the block of flats. The participant/ administrator had to decide which items belonged to whom on the basis of information s/he had about the tenants' habits (e.g. "The lamp is Mike's because he work at night"). The researcher/ assistant would return these items to the tenants. The learners were provided with pictures displaying the items and a table showing pictures of the tenants and their habits. For the information transmission group, the context of the first task was the same; however, the role of the students was different. In particular, each student was the assistant of the administrator/ researcher and s/he had to give information about what the tenants usually did at the weekend so that the administrator could make decisions about whom the owners of the items were. Hence, the information transmission group only used a table showing the residents and their habits,[5] they were not provided with pictures of the items.

In the second task of the decision-making group, each participant worked at an airport and s/he had to find the owners of lost items. S/he had to justify

[5] The table that showed pictures with habits was the same for both the decision-making and the information transmission groups in treatment tasks 1 and 2.

his/her decisions by using information about the owners' habits from a questionnaire they had filled in. Similar to task 1, the students had to use pictures of items and a table showing the owners and their habits. In the second task of the information transmission group, the context was the same as that in the second decision-making task; however, the role of the students was different. In particular, each student was an assistant who worked at the airport and in order to help the researcher, who also worked at the airport, find the owners of lost items, s/he had to give information about the habits of the owners by using a table with pictures. The treatment tasks are presented in Appendix B and their translations are given below. The prompts for all tasks were delivered in the learners' L1, given the relatively low proficiency of the learners.

(3.2)

DECISION-MAKING TASKS

TASK 1

The characters in table 1 recently moved to a new block of flats. The transportation company left their items at the entrance to the building. You [the student] are the administrator of the building and you want Nektaria, who is your assistant, to return the items to their owners. Unfortunately, you don't have their phone numbers to ask them what they have lost. However, someone who knows them gave you information about what they usually do at the weekend and you are going to use this information from table 1 to find the owners of the items. Decide which items belong to whom and justify your answers to Nektaria by mentioning what they usually do at the weekend. Mention the time and whether they are at home so that Nektaria knows whether she will find them there.

TASK 2

The characters in table 2 are tourists who lost some items at the airport. You [the student] work at the airport and you want Nektaria, who is your assistant, to return the items to their owners. However, the list of the owners of the items has been lost. Fortunately, you have information about what these people usually do at the weekend from a questionnaire they filled in. Use the information in table 2 and decide which items belong to whom. Justify your answers to Nektaria by mentioning what they usually do at the weekend. Mention the time and whether they are at home so that Nektaria know whether she will find them there.

(3.3)

INFORMATION TRANSMISSION TASKS

TASK 1

The characters in table 1 recently moved to a new block of flats. The transportation company left their items at the entrance to the building. Nektaria is the administrator of the building and she has been asked to return the items to their owners. Unfortunately, she doesn't have their phone numbers to ask them what they have lost. You [the student] know these people and what they usually do at the weekend. Nektaria needs this information to find their items. Give the information to Nektaria from table 1. Mention the activity, the time and whether they are at home so that Nektaria knows whether she will find them there.

TASK 2

The characters in table 2 are tourists who lost some items at the airport. Nektaria works at the airport and she has been asked to return the items to their owners. You [the student] have information about what these people usually do at the weekend from a questionnaire they filled in. Nektaria needs this information to find the owners of the items. Give the information to Nektaria from table 2. Mention the activity, the time and whether they are at home so that Nektaria know whether she will find them there.

For the decision-making group, the researcher did not present all of the items simultaneously so that the participants could not plan their output. The learners were only provided with one item, and after making a decision about whom the owner was, they were asked about another item. Furthermore, the items were presented in the same order to all participants. If the students of either group did not have sufficient vocabulary relating to the activities/ habits shown in the pictures, they asked in Greek and the researcher provided them with the base forms of verbs.

When the two tasks were compared in terms of their complexity, the decision-making tasks were judged to be cognitively more demanding than the information transmission ones, because they required the learners to relate information about habits to different items in order to make decisions about their owners. In contrast, the information transmission group only had to produce static information about the habits depicted in pictures. In order to validate whether the decision-making tasks imposed greater cognitive demands than the information transmission tasks, a post-task questionnaire was administered, which is described in the following section.

3.4.5 Post-task Questionnaire

After each treatment task, both groups were administered a post-task questionnaire in order to gain information related to the learners' perceptions of the tasks. Specifically, the questionnaire asked the students to judge the tasks in terms of perceived (1) cognitive demands (i.e. the mental effort they believed they had to make in order to complete the task), (2) general difficulty and (3) linguistic demands, and finally, (4) they were asked to judge the quality of their performance (i.e. whether they thought they did well or not). These judgements were elicited by a scale ranging from 1 to 9, where 1 corresponded to the lowest rate for each construct (e.g. the least mental effort), while 9 was the highest one (e.g. the greatest mental effort). The prompts were given in the students' L1. A translated version of the questionnaire is available in Appendix C.

The aim of the post-task questionnaire was to validate the construct of task complexity (i.e. whether the decision-making tasks required more mental effort than the information transmission tasks, as the researcher intended). Moreover, mental effort and task difficulty were differentiated because although a task could be perceived by the learners as relatively easy to complete, it might have still imposed greater cognitive demands due to the amount of thinking it required (see Révész, Michel, & Gilabert, 2016). For example, the participants in the decision-making condition may have easily found the owners of different items; nonetheless, they had to think in order to make connections between characters' habits and items, as opposed to the information transmission condition where the students were required to provide static information.

3.4.6 Recasts

During the treatment tasks, the learners received a recast when they produced the present third person singular inaccurately, as shown in examples 3.4 and 3.5, taken from the current study:

(3.4)

S. George is at home at 8 to 10 o'clock on Sundays and she watch a football.

R. He watches? [Recast]

S. Yes.

(3.5)

S. Because he is making a project with flowers, glue and probably scissors.

R. She makes? [Recast]

S. Yeah.

Based on Sheen (2006) and Lyster (1998), recasts in the current study had the following features: First, with regard to their *mode* and *scope*, they were interrogative and isolated as they reformulated only the erroneous part of learners' utterances without providing additional information (Lyster, 1998).

The recasts were also *reduced,* i.e. they were shorter than the learners' output (examples 3.4 and 3.5). In particular, they consisted of a personal pronoun and a verb in the present third person singular. Following Philp (2003), recasts that include fewer than five morphemes can be considered short. Thus, the recasts in the current study were classified as short.

Second, with respect to the *number of changes*, the researcher mainly provided *one-change recasts* (example 3.4) as they targeted only a single error (i.e. the present third person singular). None of the recasts of the study involved a complete rephrasing of the learners' output. However, a small number of recasts involved two changes when the personal pronoun was produced inaccurately (e.g. *she* instead of *he* followed by the present third person singular, as in example 3.5). Those recasts involving two changes were not considered distractive because the second change was always the same (i.e. the personal pronoun) and the two treatment tasks elicited frequent production of the present third person singular so as to draw the learners' attention to the tense rather than to the personal pronoun. Moreover, the personal pronoun had to be included in the input of the CF so that the learners received a meaningful obligatory context for the target construction. When only the personal pronoun was produced erroneously and the present third person singular was correct, the learners did not receive CF. Previous literature has also shown that the number of changes affects noticing only when the recasts are very different from the learners' utterance as in complete rephrasings. For example, Philp (2003) found that the number of changes of recasts negatively affected learners' noticing only in those episodes where recasts modified more than two errors in a learner's utterance.

With regard to the *type of change* made to learners' utterances, drawing on Sheen's (2006) classification, the present recasts involved either *addition* of the allomorphs /s/, /z/, /ɪz/ or /əz/ when only the base form was produced by the learner (example 3.4), or *substitution* when the researcher replaced a non-target-like utterance (e.g. another tense) with the present third person singular (example 3.5).

Finally, the *linguistic focus* of the current recasts was grammar and, specifically, a single type of grammatical error. Thus, recasts were delivered in an intensive and focused manner; however, they were elicited in a natural way as the learners performed tasks that involved a real-life context. Syntactic, lexical, pronunciation and other grammatical errors were not addressed. For example, if a learner made both a grammatical error related to the present third person singular and a lexical error associated with the choice of verb (example 3.6 below), the researcher corrected only the present third person singular and repeated the same verb as the participant. The rationale for ignoring lexical errors (e.g. "write Maths") was that the participants should not differ with respect to the type of feedback they received (i.e. whether only one error or more than one error was targeted) so that the processing demands of the feedback were similar for both groups.

(3.6)

S. Chris on Saturdays 10 and or 10 11 o'clock at home and write your Maths.

R. He writes? [Recast]

S. (Nodding)

From the data of the present study

Overall, the recasts in the current study lay towards the explicit end of the implicit/ explicit continuum due to some of the features described above (e.g. they were isolated, reduced, targeting a single error in a consistent manner). Another factor that constituted the recasts as explicit is the position of the target feature. Bardovi-Harlig (1987) has shown that when non-salient elements are placed at the end of utterances, their salience increases. In the present study, the non-salient allomorphs of the present third person singular were always at the end of recasts. Such a position might have increased the salience of the target allomorphs and facilitated their processing.

The rationale for utilizing explicit recasts was twofold. First, as noted earlier, previous studies have demonstrated that implicit recasts were less successful in promoting noticing and learning, especially when non-salient constructions were addressed (e.g. Yang & Lyster, 2010; Yilmaz, 2012). In the current study, it was expected that the salience of the present third person singular would increase as a result of the explicitness of the recasts. Second, the role of explicit recasts in facilitating noticing and interlanguage development has not received sufficient attention in previous studies.

3.4.7 Outcome measures

The study utilized a pretest and a posttest in order to gauge L2 development in knowledge of the present third person singular. The test sessions included three outcome measures: an oral production test, an elicited imitation (EI) test and a written production test. Each of these assessment tools is described below.

Oral Production Test

The aim of the oral production test was to assess the learners' ability to produce the target construction in the oral mode. The test had 12 pictures that prompted the participants to talk about the habits of fictional characters. Each of these pictures showed a different character. The pictures utilized in the pretest and posttest sessions were different; however, they depicted the same actions in order to elicit the same verbs from the students. Moreover, the number of obligatory contexts for the present third person singular was the same for both versions (i.e. pretest and posttest versions), and all of the allomorphs were produced at least once. In order to avoid any test effects, the sequence of oral production pretest and posttest was counterbalanced. The prompts for the tests were delivered in the participants' L1 (see Appendix D).

Written Production Test

The purpose of the written production test was to measure the learners' ability to produce the target feature in written mode. Similar to the oral production test, the two versions of the written test (i.e. pretest and posttest) prompted (1) the same number of obligatory contexts of the target feature (i.e. the habits of 12 characters), (2) use of the three allomorphs of the target construction at least once and (3) the two versions were counterbalanced in order to circumvent possible test effects. It should also be noted that the pictures in the written test showed the same habits as the ones in the oral production test in order to elicit the same verbs in both modes. The prompts for the written production test were the same as those for the oral production test (see Appendix E), and they were presented in the participants' L1.

Elicited Imitation (EI) Test

EI tests involve both the processing of input containing predetermined linguistic constructions and the production of output. In particular, they provide learners with grammatical and ungrammatical sentences, along with target linguistic features, and elicit their production. The question that arises is whether learners' accurate production entails rote repetition of a linguistic target or it is a result of interlanguage development. There is agreement among SLA researchers that EI tests do not simply involve the mimicking of linguistic features; on the contrary, they are inherently reconstructive (e.g.

Ellis, 2005a; Erlam, 2006; Suzuki & DeKeyser, 2015). Erlam (2006) has given several reasons for this argument. First, correcting linguistic constructions in ungrammatical sentences indicates changes in the learners' interlanguage system. If learners' accurate production was part of mere parroting, they would simply repeat the stimulus erroneously in the same way it was presented. Second, Erlam points out that, during EI tests, learners are more likely to produce linguistic elements that are already part of their interlanguage system and thus recalled from long-term memory. Research on working memory has revealed that its capacity is associated with the amount of pre-existing knowledge already stored in long-term memory (see Baddeley, Gathercole, & Papagno, 1998). The facilitative role of long-term memory has also been demonstrated in experiments whereby it was more difficult for participants to accurately repeat nonwords that were very different from real words as compared to nonwords that were similar to existing lexical items (Gathercole & Baddeley, 1993). Third, Erlam (2006) has argued that students are more likely to remember the meaning rather than the form of a sentence. Likewise, Potter and Lombardi (1990) claim that "a sentence is regenerated in immediate recall from a representation of its meaning, using recently activated words" (p. 633). To confirm this hypothesis in their experiments, Potter and Lombardi asked their participants to read or listen to a sentence (e.g. The knight rode around the palace searching for a place to enter) that contained a critical word (e.g. the word "palace"). Then, the researchers provided the participants with a list of nouns, with one of them being a synonym of the critical word (e.g. castle).

The study found that when the participants recalled the sentence, they frequently replaced the critical words with their synonyms. Interestingly, the same finding emerged for both adults and 4-year-old children.

Nonetheless, some factors that might influence the extent to which participants retain a form in their memory might be related to (1) its position in a sentence (i.e. they might be more likely to remember a linguistic construction found at the beginning or end of a sentence rather than in the middle) (Erlam & Akakura, 2016; Gallimore & Tharp, 1981) and (2) whether participants engage in tasks that direct their attention to the language or the content of the input. For example, Murphy and Shapiro (1994) demonstrated that learners were more successful in remembering the language used in the input after performing a task designed to draw their attention to linguistic elements and not to content. Hence, a reconstructive EI test that measures interlanguage development rather than rote repetition of form should encourage learners to allocate more attention to meaning rather than form by requiring them to process the semantic content of the stimulus first, before producing a sentence. To this end, in many studies utilizing EI tests, the production of sentences is preceded by a task that involves processing its content (i.e. learners are asked whether they agree or disagree with the semantic content of the stimulus or choose an image that correctly corresponds to its meaning) (e.g. Akakura, 2012; Bowles, 2011; Erlam, 2006; Spada, Shiu, & Tomita, 2015). Nonetheless, Erlam (2006) warns that the production of a stimulus may still implicate rote repetition when it occurs immediately after its presentation. For example, McDade, Simpson, and Lamb (1982) showed that when participants produced a sentence immediately after its presentation, they could repeat it accurately without necessarily comprehending it. However, they found that this was not possible when the response was elicited three seconds after presentation of the sentence. Thus, a reconstructive EI test is less likely to involve rote repetition when there is time interval between the provision of the stimulus and the learner’s production (Erlam, 2006).

Based on the above arguments, it appears that the production of utterances in an EI test can be used as a measure of interlanguage development, especially when ungrammatical sentences are corrected by test-takers. Furthermore, EI tests should encourage learners to process the meaning of stimuli (e.g. by selecting images relevant to the semantic content of sentences) so as to divert learners’ attention away from the language itself. Timed EI tests are also expected to tap into learners' implicit or automatized explicit knowledge (Erlam, 2006; Suzuki & DeKeyser, 2015).

All of the above issues were taken into account in the design of the EI test in the current study (see appendix F). The EI test sought to (1) tap into changes in the learners’ interlanguage system and (2) assess learners’ implicit or automatized explicit knowledge. The EI test was administered using Microsoft PowerPoint. After listening to a sentence, the students saw two pictures labelled A and B and were asked to choose the picture that was relevant to the meaning of the utterance they had heard by saying A or B out loud. Next, when the colour of the slide changed from white to brown, the learners had to produce the sentence in correct English. The pictures on the brown slide remained in order to help the students remember the content of the utterance. There was also a 4-second time interval between the presentation of the stimulus and the elicited response to ensure that the production of the sentence was not a result of rote repetition of the stimulus. Moreover, the learners had to produce the stimulus under time pressure. The pilot study indicated that 10 seconds were sufficient to allow the learners to recall the meaning without planning their answers. It should also be noted that the target feature (i.e. the present third person singular) was in the middle of the sentence. As explained earlier, such a position makes rote repetition less likely to occur in comparison to linguistic elements found at the beginning or end of a stimulus (Erlam & Akakura, 2016; Gallimore & Tharp, 1981).

The EI test started with seven practice items, which did not include the target construction. The main part of the EI test consisted of 72 sentences: 24 of them had items with the present third person singular while 48 of them served as distractors. Half of the sentences with the target feature were grammatical, whereas the other half were ungrammatical. In the ungrammatical items, the allomorphs of the present third person singular were omitted, only the base form was used. There was also equal distribution of the three allomorphs: three grammatical and three ungrammatical sentences for each of the three allomorphs /s/, /z/ and /əz/. With regard to the distractors, there were equal numbers of grammatical and ungrammatical sentences. Following Keating and Jegerski's (2014) suggestion, the items were pseudorandomized so that the same allomorphs of the present third person singular were not presented immediately, one after another (e.g. two verbs with the allomorph /s/ consecutively). They were presented on the basis of the pattern shown in Table 8.

TABLE 8 PSEUDORANDOMIZATION PATTERN OF THE EI TEST

Pretest

ITEM	VERSION A	VERSION B
Distractor	I'm reading a book now.	
Distractor	The kitchen was such small.	
Target item /s/	He always take the bus.	He always takes the bus.
Distractor	I'm open the door now.	
Distractor	She's wearing a nice hat.	
Target item /z/	He sometimes gives a sweet.	He sometimes give a sweet.
Distractor	My brother made this cake.	
Distractor	She'll cooking for a friend.	
Target item /ɪz/	He often miss the bus.	He often misses the bus.
Distractor	I wanted wear a dress.	
Distractor	Her salad was the best.	
Target item /s/	He always sleeps at home.	He always sleep at home.
Distractor	My sister found a dog.	
Distractor	The kitchen was so small.	
Target item /z/	He often cry at night.	He often cries at night.
Distractor	My sister can to drive.	
Distractor	He arrived in France late.	
Target item /ɪz/	He always uses a spoon.	He always use a spoon.

Posttest

ITEMS	VERSION A	VERSION B
Distractor	I'm writing a book now.	
Distractor	The village was such small.	
Target item /s/	He always take the train.	He always takes the train.
Distractor	I'm open the bag now.	
Distractor	She's wearing a nice skirt.	
Target item /z/	He sometimes gives a cake.	He sometimes give a cake.
Distractor	My brother made this film.	
Distractor	She'll cooking for her dad.	
Target item /ɪz/	He often miss the train.	He often misses the train.
Distractor	I wanted wear a shirt.	
Distractor	Her salad was the worst.	
Target item /s/	He always sleeps at work.	He always sleep at work.
Distractor	My sister found a cat.	
Distractor	The village was so small.	
Target item /z/	He often cry at home.	He often cries at home.
Distractor	My sister can to sing.	
Distractor	He arrived in Greece late.	
Target item /ɪz/	He always uses a glass.	He always use a glass.

Based on Keating and Jegerski (2014), in order to combat item effects, two versions were designed for each item that differed as to whether the target verb was presented as grammatical or ungrammatical. Example 3.7, below, demonstrates the two versions of the target verb "take", and table 8, above, shows both versions of some of the target items. The learners did not encounter two versions of the same item in order to avoid *repetition* *effects* that would lead to unnatural reactions during production of the second version (e.g. repeating it quickly and superficially because they have seen it before) (Keating & Jegerski, 2014). Thus, half of the students were administered version A and the other half version B, in both groups. The positions of the test items and the pretest and posttest versions were also counterbalanced.

(3.7)
Version A – ungrammatical version of the item
"He always take the bus"
Version B – grammatical version of the item
"He always takes the bus"

Drawing on Keating and Jegerski's (2014) suggestions, the critical verbs in the present third person singular were the same in length (i.e. one-syllable verbs) and always located in the same position in all the sentences in order to impose similar processing demands on the learners. Moreover, according to Vocabprofiler (available online) (Cobb, 2016; Heatley, Nation, & Coxhead, 2002), all the verbs were found to be words frequently used in English as they corresponded to either K1 or K2 categories. K1 includes the first thousand and K2 those and the second thousand most frequent lexical items in the English language. The rationale for including some K2 words was also that they are frequently presented in textbooks and the pilot study showed that they were familiar to the students (e.g. the verb "dance"). With respect to the *precritical region*, i.e. the construction prior to the present third person singular, it consisted of a personal pronoun serving as the subject of the critical verb. Human beings' names or longer noun phrases were not used as subjects to ensure that the target items did not differ with regard to the difficulties the learners might experience in processing the *precritical region*. When the processing demands of the precritical region spillover the target verb, they might affect the learners' performance during its production (Keating & Jegerski, 2014). Two-syllable adverbs of frequency were also part of the precritical region, between the personal pronoun and the target verb, so as to create an obligatory context for the present third person singular. The length of the stimuli that included the target verb and the distractors was the same. In particular, all 72 sentences consisted of six syllables. The distractors were also twice as numerous as the sentences with the critical region (i.e. the present third person singular) in order to obscure the target of the study. The reliability of the two versions of the EI test (one serving as a pre-test, the other as a post-test) was examined by calculating Cronbach's Alpha, and both versions were found to be reliable ($\alpha = .842$ for version one, $\alpha = .867$ for the other version).

3.4.8 Rationale for Test Selection

The three outcome measures presented above were selected in consideration of the following factors: (1) whether they involved production in the oral or written mode, (2) whether they were designed to tap into implicit, explicit declarative knowledge or procedural knowledge and (3) the extent to which they enabled learners to control their L2 output. Each of these factors is discussed below.

Mode

The oral production and EI tests were conducted in oral mode, as opposed to the written production test. Considering that oral recasts were delivered during the treatment, those tests that required oral production of the target construction were closer to the stimulus the learners received (i.e. oral feedback). In contrast, the written production test diverged more from the oral treatment. Hence, it provided insights into whether possible positive effects of oral CF would transfer to the written mode.

Type of Knowledge

The oral production test primarily assessed the learners' procedural knowledge of the target construction, and it enabled them to rely less on their explicit, declarative knowledge in comparison to the written production test. Although the learners performed the task at their own pace without any time limit, oral production naturally imposes greater time pressure than production in the written mode and, consequently, it implicates less use of explicit, declarative knowledge (see DeKeyser (2007) and DeKeyser & Criado (2013) for a review of skill acquisition theory).

Regarding the EI test, there is a heated debate among researchers as to whether elicited imitation serves as a measure of implicit knowledge (e.g. Bowles, 2011; Ellis, 2005a; Erlam, 2006; Erlam and Akakura, 2016; Spada, Shiu, & Tomita, 2015) or automatized explicit knowledge (e.g. Suzuki & DeKeyser, 2015). The former means that learners are not aware of the target construction when producing it during a test (see Rebuschat, 2013). Counter to this argument, Suzuki and DeKeyser (2015) found that learners' scores on a timed EI test positively correlated with their metalinguistic knowledge, while they were not related to their performance on a serial reaction time task (SRT), which is a measure of aptitude for implicit learning. Based on these results, the researchers argued that learners' successful performance on an EI test does not necessarily imply implicit knowledge (i.e. a lack of awareness during accurate production). They contend that even when time pressure is imposed, learners may still deploy their explicit knowledge when it has been automatized.

The present thesis assumed that the EI test measured implicit knowledge or automatized explicit knowledge since, drawing on Suzuki and DeKeyser's (2015) findings, the possibility that the participants were aware of the target construction while taking the test cannot be excluded. Nonetheless, it should be emphasized that the learners had only four seconds to process the meaning of the sentences they heard by choosing the correct picture, and ten seconds to produce the sentences in correct English. Considering the time pressure imposed on the participants in tandem with the fact that the test was administered in oral mode, it is presumed that the students who achieved high scores where those who were able to use explicit knowledge that had reached a stage of automatization or relied on implicit knowledge. It should also be underlined that the test-takers’ scores were based on their first answers, self-corrections were excluded.

Unlike the oral production and EI tests, the written production test in the present thesis was expected to primarily involve the use of explicit, declarative knowledge, not necessarily automatized. The learners were asked to write sentences that elicited the present third person singular without time pressure. They also had time to revise their sentences at the end of the test and make any necessary changes. In other words, the participants produced the target construction at their own pace and they could also revise their output. As no time limit was imposed, the written production test enabled the students to draw on their explicit, declarative knowledge in addition to deploying their procedural knowledge. Indeed, the presence of self-corrections during revisions suggests that the learners did rely on their explicit, declarative knowledge when completing the test.

Degree of Control over their L2 Output

The three tests also differed as to whether they allowed the learners to control their output. In particular, during the EI test, the learners' production was based on pre-determined input as they had to produce utterances provided by the test. In the oral and written production tests, the students' output was also controlled as it was prompted by pictures; however, they were not given specific lexical items to produce and could use the pictures in a flexible way by using their preferred vocabulary (e.g. She feeds her horse / she takes care of her horse/ she gives food to her horse).

3.4.9 L2 Aptitude Test

The participants were also asked to complete a computer-administered aptitude test called LLAMA (Meara, 2005). The LLAMA test consists of four components: LLAMA B measures learners' ability to learn vocabulary in a short period of time, LLAMA D assesses learners' ability to recognize patterns in spoken language, LLAMA E demonstrates the extent to which learners are capable of associating sounds with symbols (i.e. phonetic coding) and LLAMA F is a measure of learners' grammatical inferencing. The LLAMA test is language-independent as it uses visual stimuli and a language unfamiliar to the participants, a dialect of an Indian language spoken in British Columbia in Canada. The test instructions were given in the students' L1 following the LLAMA manual (Meara, 2005). The following sections explain what the participants were asked to do for each test component.

LLAMA B

LLAMA B is a vocabulary learning task. First, the learners had two minutes to click on objects and learn their names (20 in total). Then, the names of the objects appeared on screen, and the participants were asked to match the objects with their names at their own pace.

LLAMA D

LLAMA D is a sound recognition test providing insights into learners' ability to recognize patterns in spoken language. According to Meara (2005), it also measures the ability to recognize morphological variations related to grammar. First, the participants were asked to listen to a string of 10 computer-generated sound sequences only once. Then, they were administered a recognition test that required them to distinguish sounds they had already heard from novel ones, without time pressure.

LLAMA E

LLAMA E is a sound-symbol correspondence task measuring learners' phonetic coding ability. First, the learners had two minutes to listen to 24 recorded syllables provided in combination with their transliterations, and they were expected to work out the sound-symbol correspondence for each syllable. Then, they had to listen to a two-syllable word and, simultaneously, were given two possible written representations. The learners had to choose the symbols that correctly corresponded to the words they heard at their own pace.

LLAMA F

LLAMA F has been designed to measure grammatical inference. The learners were presented with 20 sentences accompanied by pictures. Each of these sentences described a picture displayed on screen. The participants had five minutes to read the sentences, look at the corresponding pictures, and figure out the grammatical rules of the new language. This task was followed by a test with items that had a picture and two sentences: one grammatical, one ungrammatical. The test-takers had to choose the correct sentence for each picture without time pressure.

3.4.10 Data Collection

The data were collected from 60 participants over a time span of six months. The researcher met each young learner and his/her parents individually, prior to the experiment, to give them an information sheet explaining the purpose of the study and answer any additional questions. If they were interested in participating, a consent form was given to the children and their parents. Then, each participant was asked to fill in a background questionnaire asking information about their age, gender, L1, years of learning English, whether they had lived in an English-speaking country prior to the study and if so for how long, how often they used a keyboard (see Appendix A). The questionnaire was in the learners' L1. On the first day of the study, the students were administered a proficiency test. When the test was completed, it was scored by the researcher and if a student's proficiency level was appropriate for the study, they continued with the pretests. They were first subjected to an oral production pretest, followed by an EI pretest and then a written production pretest. The rationale for starting with an oral test followed by a written test was to encourage the participants to produce the target feature without having sufficient time available to resort to their explicit, declarative knowledge during oral production. As for the order of the two oral tests, the oral production test preceded the EI test, so that learners produced the target construction without being exposed to any input (i.e. during the EI test they had to listen to sentences containing the present third person singular). Regarding the time to take the assessment tasks, the oral production test lasted for approximately 5 minutes, and the EI test around 25 minutes. The length of the written production test was approximately 15 minutes. After completing the pre-tests, the learners were also asked to take three components of the aptitude test (LLAMA D, E and F). As aptitude was expected to affect development in knowledge of the target feature, the researcher made sure to control for differences in aptitude across the two groups using stratified random assignment.

Those learners who had very limited prior knowledge of the target construction, as demonstrated by the three outcome measures, were invited to attend treatment sessions, which started approximately a week after the pretest. The treatment was conducted individually, and it consisted of two tasks: either two information transmission or two decision-making tasks, depending on group assignment. Both groups received recasts from the researcher in response to errors on the present third person singular. The two treatment sessions lasted for approximately 10 minutes each. Each treatment task was followed by a post-task questionnaire. After the treatment, the learners took a five-minute break and were asked to complete the last component of the aptitude test (LLAMA B). Next, they were administered a posttest that also required oral production, EI and written production of the target feature. Similar to the pretest, the posttest lasted for approximately 45 minutes. LLAMA B was administered between the treatment and posttests to function as a distraction from intensive feedback. Figure 2 shows a schematic representation of the experimental schedule.

FIGURE 2 EXPERIMENTAL SCHEDULE (STUDY 1)

SESSION 1

Consent form and background questionnaire (5 minutes)

Proficiency test (10 minutes)

Pretests

Oral production (5 minutes)

Elicited imitation (25 minutes)

Written production (15 minutes)

Aptitude test (LLAMA D, E and F)

(one-week interval)

SESSION 2

Information transmission group: 2 tasks (20 minutes)

Treatment

Decision-making group: 2 tasks (20 minutes)

(Break, approximately 5 minutes)

Aptitude (LLAMA B)

Posttests

Oral production (5 minutes)

Elicited imitation (25 minutes)

Written production (15 minutes)

3.4.11 Data Analyses

Transcription

All the oral production data from the treatment tasks and tests were transcribed by the researcher. Ten per cent of the data were also transcribed by a second researcher to verify the reliability of the transcriptions. The data transcribed by the second researcher were selected through stratified random sampling that ensured equal representation of the experimental groups and their performances across all oral tasks and tests.

The two transcripts were compared with a focus on the target verbs. Discrepancies between the two transcripts referred to (1) items that were differently transcribed and (2) items present in one transcript but omitted from the other one. Inter-transcriber agreement was calculated by dividing the total number of items transcribed identically by the total number of items, and it was found to be high (.988). Cohen's kappa was also computed and reached .962.

Coding and Scoring of Uptake with Output Modification

The learners in both groups were always given an opportunity to modify their output as there was no topic continuation after the provision of recasts. Drawing on previous literature, uptake was coded as successful when learners modified their output and corrected their initial errors on the present third person singular by producing the same verb (example 3.8) or a new verb (example 3.9). In contrast, their uptake was coded as unsuccessful when they (1) repeated the same error by producing the present third person singular inaccurately on the same or a different verb (example 3.10), (2) when they made an error related to a different construction without correcting the present third person singular (example 3.11), (3) when they gave a yes/no answer (examples 3.12 and 3.13), (4) when they remained silent (example 3.14) and when they continued to give information (topic continuation as in the example 3.15) (for a review of the coding used, see Lyster & Ranta, 1997). A more general term, modified output, rather than uptake is used in the current thesis as it involves both repetitions of recasts (i.e. production of the same verb as shown in example 3.8) or application of the present third person singular pattern in a new verb after receiving a recast (example 3.9).

(3.8)

S. Helen is at home at 9 to 11 11 o'clock on Saturdays and she watching TV

R. She watches?

S. Ναι (yes) she watches. (successful uptake – modified output, same verb)

(3.9)

S. He she is listening to music?

R. She listens?

S. No. She sings. (successful uptake – modified output, new verb)

(3.10)

S. Helen on Saturday at 9 11 p.m. watch television

R. She watches?

S. She watch television at home. (No modified output – same error)

(3.11)

S. Jessica on Saturday on Saturdays stays at home and is makes pictures

R. She makes?

S. Is drawing (No modified output – different error)

(3.12)

S. Jessica because the the draw a woman at home.

R. She draws?

S. Ναι (Yes). (No modified output – answering yes)

(3.13)

S. The map maybe belongs to Mary because on Sundays at all day he is not at home and he is climbing on the mountain.

R. She climbs?

S. Not at all. (No modified output – answering no)

(3.14)

S. Alice on Saturdays 10 or 12 year 12 o'clock is at home and make your food.

R. She makes?

S. (Nodding) (No modified output – silence and nodding)

(3.15)

S. Jack on Saturdays 9–11 at home and play computer games.

R. He plays?

S. Natalie on Sundays 3–5 at home and reading a book read a book. (No modified output – topic continuation)

Coding and Scoring of the Elicited Imitation Test

The scoring criteria for the EI test were based on Erlam (2006). The learners were given two scores for those items that included the present third person singular: one score reflecting their understanding of the meaning of the sentence, and one capturing the correct use of the form when producing the sentence. With regard to meaning, the learners were given 1 point if they selected the correct picture and 0 points if they did not. This score was calculated in order to test whether the learners had processed the semantic content of the utterance. Only students who chose the correct picture at least 90 per cent of the time were included in the study. Regarding their scores for form, in line with Erlam (2006), the coding was based on the three categories described below:

1. *Obligatory occasion created*: <u>supplied</u> [1 point] (i.e. An obligatory context was created in the learner's utterance and the target construction was produced correctly. Errors on vocabulary or other errors related to grammar were ignored as they were not the focus of the study) (example 3.16).

2. *Obligatory occasion created*: <u>not supplied</u> [0 points] (i.e. An obligatory context was created in the learner's utterance but the target feature was produced incorrectly) (examples 3.17 and 3.18).

3. *No obligatory occasion created* [0 points] (i.e. The learner's response did not create an obligatory context that required use of the target feature even if they had to repeat an utterance that prompted its use. For example, when the learner produced a different linguistic construction, their response was considered as avoidance of the target feature) (example 3.19).

The examples below show how the coding was realized in the data.

(3.16)

R. He sometimes help his mum.

S. A

S. He sometimes *helps* his mum. [1 point] (Obligatory context created – supplied)

(3.17)

R. He sometimes watch a match.

S. A

S. He sometimes *watching* a match. [0 points] (Obligatory context created – not supplied, different construction used)

(3.18)

R. He sometimes gives a cake.

S. B

S. He sometimes *give* the cake. [0 points] (Obligatory context created – not supplied, only base form used)

(3.19)

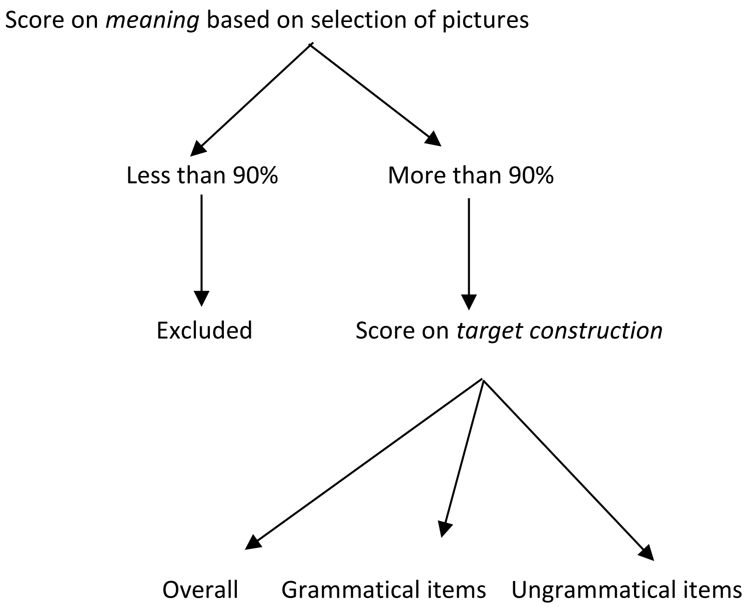
R. She always dances with George.

S. A

S. She is *dancing* with George the George. [0 points] (Obligatory context created – avoidance, production of a different linguistic construction).

Using the categories presented above, the researcher calculated (1) the overall score for production of the present third person singular for the grammatical and ungrammatical items combined, (2) the score for grammatical items and (3) the score for ungrammatical items. Figure 3 demonstrates the coding process for the EI test. The gain score was calculated by subtracting the pre-test score from the post-test score.

FIGURE 3 EI CODING



<u>Coding and Scoring of the Oral and Written Production Tests</u>

The data from the oral and written production tests were coded by following the two main steps demonstrated in Figure 4. First, obligatory contexts (OCs) for the present third person singular were identified. On the oral production test, repetitions were coded as new OCs when they involved (1) two or more erroneous productions (sometimes of the same verb) (example 3.20), (B) self-corrections (example 3.21), (C) accurate production of a verb followed by an erroneous one (either of the same or a different verb) (example 3.22). On the written production test, there were no such repetitions.

(3.20)

Nick on Saturday he *make* (OC1) he *make* (OC2) a snowman.

(3.21)

Nick *make* (OC1) *makes* (OC2) a snowman.

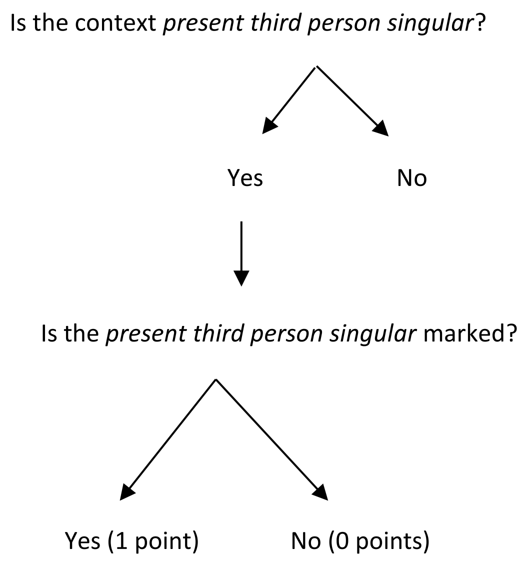
(3.22)

She he *builds* (OC1) *make* (OC2) a snowman.

After identification of the OCs, the following coding scheme was used: If the target feature had been accurately produced, the students received 1 point. Vocabulary errors were ignored. That is, when participants produced the wrong verb but had used the present third person singular accurately (e.g. "He makes the car" instead of "he fixes the car"), they received 1 point, as they exhibited development in the target feature. When the base form or another tense was produced, then the score was 0 points. When the learners added one of the three allomorphs /s/, /z/, /ɪz/ or /əz/ but in combination with auxiliary verbs used in different tenses (e.g. is/was runs), 0 points were given to the student. Pronunciation errors were ignored. For example, regarding verbs such as “plays”, if a student incorrectly used the allomorph /s/ as in [pleɪs] instead of /z/ as in [pleɪz], this error was ignored as the focus of the study was on the development of morphology and not pronunciation. Similarly, in the production of verbs with the allomorph /əz/ as in the verb “watches”, if a participant used the sound /es/ as in [wɒtses] instead of /əz/ as in [wɒʧəz], this was not taken into account. Replacing /əz/ with /es/ is a very common pronunciation error for Greek learners of English as a schwa vowel does not exist in Greek. Nonetheless, the aim of the feedback was not to address such pronunciation errors, and hence they did not influence scores on the oral tests (i.e. neither the oral production test nor the EI test).

On the written production test, when the students produced one of the written symbols –s, -es, -ies erroneously (e.g. He fixs), it was coded as correct and 1 point was given. The rationale for ignoring spelling errors was that spelling was not the focus of the study, and the participants received oral feedback, which means that they were not expected to gain any knowledge on written alternations –s, -es, -ies. For example, if a learner produced the sentence “He fix” on a written pretest, and “He fixs” on a written posttest, the oral feedback was successful in terms of inducing morphological development despite the error in spelling. Similar to the oral production test, the students received 1 point when the present third person singular was produced correctly, even on an erroneous verb. In contrast, when the students wrote only the base form of the verb, or a different tense, they received 0 points. On both the oral and written production tests, learners’ gains were computed by subtracting their pretest score from their posttest score.

FIGURE 4 ORAL AND WRITTEN PRODUCTION TESTS CODING



<u>Inter-coder Agreement</u>

Ten per cent of the data were coded by a second researcher (both modified output in the treatment tasks and target verbs on the tests). There was equal representation of the experimental groups. According to Table 9, there was a high level of agreement between the two coders (.98). The Cohen’s kappa values were also high: .93 for modified output produced in the treatment tasks, .95 for the oral production test, .96 for the written production test and .97 for the EI test.

TABLE 9 INTER-CODER AGREEMENT AND COHEN'S KAPPA

	Inter-coder agreement	Cohen's kappa
Modified output	.98	.93
Oral production test	.98	.95
Written production test	.98	.96
EI test	.98	.97

3.4.12 Results

Preliminary Analyses

The current section presents the preliminary statistical analyses conducted prior to exploring the research questions. Such preliminary analyses sought to investigate (1) whether the participants' gain scores were normally distributed so as to use appropriate statistical tests, (2) whether the decision-making tasks were perceived as more complex than the information transmission tasks, based on learners' answers in the post-task questionnaire and (3) if there were any significant differences between the two groups in their pre-test performance, in their aptitude scores and in the opportunities they had to receive recasts. The latter was examined by ensuring that the information transmission and decision-making tasks involved a similar number of obligatory contexts (OCs) for the target construction. Each of these steps is discussed in detail below.

Before delving into the research questions of the thesis, a Shapiro-Wilk test, which is a test of normality, was run in order to examine whether the participants' gain scores were normally distributed. The test reached significant levels (Table 10), indicating that non-parametric statistical procedures were required. Hence, the current study employed the Mann-Whitney U test, and the effect sizes were calculated by using the formula $r=z/\sqrt{N}$. The interpretation of the effect sizes was based on Plonsky and Oswald's (2014) suggestion that *r* close to .25 is small, .40 medium and .60 large.

TABLE 10 TEST OF NORMALITY (SHAPIRO-WILK)

TYPE OF ASSESSMENT TASK	N	GROUP	Shapiro-Wilk (Sig.)
Oral production test	30	Information transmission	.002
	30	Decision-making	.001
Written production test	30	Information transmission	< .001
	30	Decision-making	< .001
EI test (overall)	30	Information transmission	.017
	30	Decision-making	.012
EI test (grammatical)	30	Information transmission	.069
	30	Decision-making	.070
EI test (ungrammatical)	30	Information transmission	< .001
		Decision-making	< .001

Next, in order to validate the construct of task complexity, a Mann-Whitney U test, i.e. a non-parametric test, was run based on the post-task questionnaire data. Tables 11 and 12 show that the students judged the decision-making tasks to be more cognitively demanding than the information transmission tasks, as intended by the researcher. Specifically, the mean values for mental effort were 5.93 and 5.23 for the two decision-making tasks, whereas they were 3.83 and 3.43 for the two information transmission tasks. As Table 13 indicates, this difference was significant for both tasks (Mann-Whitney U = 191.00, $p < .001$, and Mann-Whitney U =251.50, $p = .003$ for tasks 1 and 2, respectively). The effect size was large for task 1 ($r = .50$) and medium for task 2 ($r = .38$). Thus, the construct of task complexity was validated by the mental effort scale as the participants perceived the decision-making tasks to require more mental effort than the information transmission tasks. No differences were found between the complex and simple conditions in the participants' perceptions related to overall difficulty, linguistic difficulty and performance (see Tables 11, 12 and 13).

TABLE 11 DESCRIPTIVE STATISTICS FOR TASK 1 PERCEPTIONS PER GROUP

TYPE OF ASSESSMENT TASK	*GROUP (N=30)*	*Mean*	*SD*	*Median*	*IQR*
Mental effort	Information transmission	3.83	1.74	3.50	3.00
	Decision-making	5.93	1.91	6.00	2.25

Task Overall Difficulty	Information transmission	2.53	1.59	2.00	3.00
	Decision-making	2.63	1.67	2.00	3.00
Linguistic Difficulty	Information transmission	3.30	1.70	3.00	2.25
	Decision-making	3.20	1.62	3.00	2.25
Performance	Information transmission	3.87	1.94	4.00	3.00
	Decision-making	3.27	2.05	3.00	3.00

TABLE 12 DESCRIPTIVE STATISTICS FOR TASK 2 PERCEPTIONS PER GROUP

TYPE OF ASSESSMENT TASK	*GROUP (N=30)*	*Mean*	*SD*	*Median*	*IQR*
Mental effort	Information transmission	3.43	1.71	3.00	3.00
	Decision-making	5.23	2.35	5.50	4.00
Task Overall Difficulty	Information transmission	2.57	1.30	2.00	2.25
	Decision-making	2.60	1.54	2.00	2.25
Linguistic Difficulty	Information transmission	3.13	1.87	3.00	2.00
	Decision-making	2.73	1.59	2.00	3.00
Performance	Information transmission	3.67	1.86	3.50	3.00
	Decision-making	3.43	2.19	3.00	3.00

TABLE 13 BETWEEN GROUP DIFFERENCES IN LEARNERS' TASK PERCEPTIONS

TYPE OF ASSESSMENT TASK	*TASK 1*			*TASK 2*		
	Mann-Whitney U	*Sig. p*	*Effect size r*	*Mann-Whitney U*	*Sig. p*	*Effect size r*
Mental effort	191.00	<.001	.50	251.50	.003	.38
Task Overall Difficulty	435.50	.824	.02	441.00	.891	.01
Linguistic Difficulty	437.00	.843	.02	399.00	.441	.09
Performance	362.00	.186	.17	399.00	.445	.09

As explained earlier, it was also explored whether the two groups were comparable. To this end, a series of Mann- Whitney U tests were run to examine (1) whether there were any significant differences between the two experimental groups on any of the pretests, (2) whether there were any significant differences in the numbers of OCs created by the information transmission and decision-making tasks during the treatment so as to confirm that the two groups had equal opportunities to receive recasts and (3) whether the two groups differed in terms of their aptitude scores.

First, in order to ensure that potential differences between the two conditions in the gains exhibited after the treatment can be attributed to the combined effects of task complexity and recasts rather than to learners' prior knowledge of the target feature, the researcher examined whether the two groups were similar in their pretest scores. Tables 14 and 15 show descriptive statistics for the learners' pretest performances on oral production, written production and elicited imitation. As shown in Table 16, there were no significant differences between the two groups on any of the pretests (p > .05). The effect sizes were small (r = < .01 for the oral production pretest, r = .04 for the written production pretest, r = .03 for the EI pretest, r = .03 for the EI grammatical items, r = .08 for the EI ungrammatical items).

TABLE 14 DESCRIPTIVE STATISTICS FOR THE ORAL AND WRITTEN PRODUCTION PRETEST SCORES OF THE TWO GROUPS

TYPE OF ASSESSMENT TASK	*GROUP*	*N*	*Mean (%)*	*SD*	*Median*	*IQR*
Oral production test	Information transmission	30	3.99	8.19	.00	5.73
	Decision-making	30	3.75	6.91	.00	6.08
Written production test	Information transmission	30	4.10	9.22	.00	2.08
	Decision-making	30	4.93	9.30	.00	8.33

TABLE 15 DESCRIPTIVE STATISTICS FOR THE EI PRETEST SCORES OF THE TWO GROUPS

TYPE OF ASSESSMENT TASK	*GROUP*	*N*	*Mean*	*SD*	*Median*	*IQR*
EI test (overall)	Information transmission	30	4.26	4.25	3.00	4.50
	Decision-M making	30	4.00	4.10	3.00	6.00
EI test (grammatical)	Information transmission	30	3.56	3.51	3.00	4.00
	Decision-making	30	3.50	3.63	2.00	6.00
EI test (ungrammatical)	Information transmission	30	.70	1.14	.00	1.00

Decision-making	30	.50	.86	.00	1.00

Note: The total score was 24 for the EI test (12 for grammatical items, 12 for ungrammatical items).

TABLE 16 BETWEEN GROUP DIFFERENCES ON THEIR PRETEST SCORES

TYPE OF ASSESSMENT TASK	*Mann-Whitney U*	Sig. p	*Effect size r*
Oral production pretest	448.00	.970	<.01
Written production pretest	433.00	.741	.04
EI pretest (overall)	430.50	.771	.03
EI pretest (grammatical)	432.50	.794	.03
EI pretest (ungrammatical)	414.00	.534	.08

Second, the researcher investigated whether the two conditions (simple vs complex) were similar with respect to the number of OCs of the target feature to ensure the two groups had equal opportunities to receive recasts. Table 17 presents descriptive statistics. As shown in Table 18, there were no significant differences in the numbers of OCs between the two groups for either of the treatment sessions ($p > .05$). In other words, the learners had comparable opportunities to be provided with recasts and, consequently, any differences in gains between the two conditions could be attributed to the combined effects of task complexity and recasts rather than to differential amounts of CF supplied. The effect size for the differences in OCs produced in the treatment sessions was small ($r = .17$).

Table 17 Descriptive Statistics for the Amount of OCs per Group

GROUP	*Number*	*Mean (%)*	*SD*	*Median*	*IQR*
INFORMATION TRANSMISSION (TASK 1)	458	15.26	2.95	15.00	1.25
INFORMATION TRANSMISSION (TASK 2)	462	15.40	2.19	15.00	2.00
INFORMATION TRANSMISSION TOTAL	920	30.66	5.01	30.00	2.50
DECISION-MAKING (TASK 1)	489	16.30	4.13	15.00	2.25
DECISION-MAKING (TASK 2)	491	16.36	3.89	15.00	1.50
DECISION-MAKING TOTAL	980	32.66	7.71	30.00	2.25

Table 18 Between Group Differences in the Amount of OCs

Treatment	*Amount of OC*		
	Mann-Whitney U	*Sig. p*	*Effect size r*
TASK 1	362.00	.181	.17
TASK 2	398.00	.420	.10
BOTH TASKS	359.50	.172	.17

Finally, it was examined whether there were any significant differences between the two groups in terms of L2 aptitude. Table 19 presents descriptive statistics and Table 20 indicates that there were no significant differences between the two groups of participants regarding their aptitude scores (p >.05). The effect sizes were r =.07 for LLAMA B, r =.08 for LLAMA D, r =.03 for LLAMA E and r = <.01 for LLAMA F.

Table 19 Descriptive Statistics for the L2 Aptitude Scores per Group

LLAMA COMPONENT	*GROUP*	*N*	*Mean*	*SD*	*Median*	*IQR*
LLAMA B	Information transmission	30	32.50	16.90	30.00	21.25
	Decision-making	30	28.66	13.95	25.00	25.00
LLAMA D	Information transmission	30	23.66	11.36	25.00	15.00
	Decision-making	30	22.16	16.33	20.00	26.25
LLAMA E	Information transmission	30	42.66	26.25	40.00	30.00
	Decision-making	30	43.00	31.85	50.00	60.00
LLAMA F	Information transmission	30	21.66	20.35	20.00	40.00
	Decision-making	30	21.66	19.84	20.00	32.50

Table 20 Between Group Differences in L2 Aptitude

LLAMA COMPONENT	*Mann-Whitney U*	*Sig. p*	*Effect size r*
LLAMA B	409.00	.541	.07
LLAMA D	407.50	.527	.08
LLAMA E	433.00	.800	.03
LLAMA F	447.00	.964	< .01

Overall, the two groups were similar with regard to their prior knowledge of the target construction, the numbers of opportunities they had to receive recasts (i.e. number of OCs of the present third person singular) and their L2 aptitude. After confirming that the two groups were comparable in terms of these factors, another series of Mann-Whitney U tests were run in order to answer two of the research questions: (1) whether there were any significant differences between the two groups in the amount of modified output following recasts and (2) whether there were any significant differences between the two groups with respect to their gain scores on each test (i.e. oral production, written production, EI test). As explained earlier, following Plonsky and Oswald (2014), effect-size estimates were calculated by using the formula $r = z/\sqrt{N}$ and values close to .25, .40 and .60 were interpreted as small, medium and large, respectively.

Regarding research questions that involved correlations, Spearman tests were run to investigate whether there was a relationship between (1) the amount of modified output and subsequent L2 development, (2) aptitude and modified output regardless of task complexity, (3) aptitude and L2 development irrespective of task complexity, (4) aptitude and modified output for each group separately and (5) aptitude and L2 development for each experimental group. The results for each research question are presented in the following section.

Research Question 1

The first research question addressed whether and to what extent task complexity affected the amount of output modification produced by L2 learners after the provision of recasts targeting the present third person singular. Table 21 provides descriptive statistics for the amount of output modification per group. As the table shows, there were fewer than 30 students who modified their output in the second treatment task of each condition (N = 27 in the information transmission group, N = 29 in the decision-making group), because some participants managed to produce the target feature accurately during these tasks, and hence no recasts were supplied. The table also demonstrates that the information transmission or low-task complexity group modified their output to a greater extent than the decision-making or high-complexity group. In particular, the mean percentage of output modification for the simple condition was 51.45%, whereas for the complex condition it was 40.50%. The difference, however, between the two groups in the amount of modified output was not significant (Mann-Whitney U = 388.50, p = .362) (Table 22). The effect sizes were also small (r =.12 for task 1, r = .13 for task 2, r =.11 for both tasks). Hence, the two groups generated similar amounts of modified output regardless of whether they received recasts during complex or simple tasks. It should also be noted that the interquartile range (IQR) was large for both groups, indicating great variability.

TABLE 21 DESCRIPTIVE STATISTICS FOR OUTPUT MODIFICATION PER GROUP

	GROUP	*N*	*Mean (%)*	*SD*	*Median*	*IQR*
INFORMATION TRANSMISSION	TASK 1	30	49.84	36.57	53.84	66.67
	TASK 2	27	56.02	43.37	66.66	100.00
	TOTAL	30	51.45	37.10	62.50	75.00
DECISION-MAKING	TASK 1	30	38.89	34.61	33.33	61.54
	TASK 2	29	43.39	40.56	44.44	90.00
	TOTAL	30	40.50	35.16	37.50	57.82

TABLE 22 BETWEEN GROUP DIFFERENCES IN OUTPUT MODIFICATION

GROUP	*Mann-Whitney U*	*Sig. p*	*Effect size r*
TASK 1	383.50	.324	.12
TASK 2	330.50	.304	.13
TOTAL	388.50	.362	.11

Research question 2

The second research question investigated the combined effects of task complexity and recasts on promoting knowledge of the present third person singular. The descriptive statistics are presented in Tables 23 and 24. As Table 23 indicates, on the oral and written production tests, both groups improved their scores; however, the information-transmission group outperformed the decision-making group. In particular, on the oral production test, the students in the information transmission condition benefitted from the treatment twice as much as the decision-making condition (i.e. the gain scores were 29.61% and 13.55% for the two groups, respectively). On the written production test, the information transmission group's gain score reached 52.94%, as opposed to the decision-making group that demonstrated lower gains (only 21.83%). As for the EI test, which consisted of 24 items overall, Table 24 shows that the learners achieved the least improvement (i.e. 2.53 for the information transmission condition, 2.00 for the decision-making condition). Not surprisingly, both groups achieved greater improvement on grammatical EI items than on ungrammatical ones.

As shown in Table 25, only on the oral and written production tests did these differences prove significant, with the information transmission group exhibiting significantly greater gains than the decision-making group (Mann-Whitney U = 311.50, p = .037 on the oral production test, Mann-Whitney U = 278.00, p = .009 on the written production test). The effect size was medium for the written production test (r = .33). As for the other outcome measures, the effect sizes were small (r = .26 for oral production gain scores, r = .07 for EI gain scores, r =.12 for EI grammatical gain scores, r =.05 for EI ungrammatical gain scores). Thus, receiving recasts in the simple condition benefited learners to a greater degree than in the complex condition. This difference was significant only when L2 development was measured by an oral and a written production test; however, the effect size was bigger on the written production test in comparison to the oral production test. Regarding the groups' IQR, greater variability was found in the simple condition (i.e. information transmission group) as compared to the complex condition (i.e. decision-making group) on the oral and written production tests.

TABLE 23 DESCRIPTIVE STATISTICS FOR GAIN SCORES ON THE ORAL AND WRITTEN PRODUCTION TESTS PER GROUP

TYPE OF ASSESSMENT TASK	*GROUP*	*N*	*Mean (%)*	*SD*	*Median*	*IQR*
Oral production test	Information transmission	30	29.61	28.91	31.16	57.19
	Decision-making	30	13.55	24.64	3.33	23.16
Written production test	Information transmission	30	52.94	45.24	63.72	100.00
	Decision-making	30	21.83	38.35	.00	52.08

TABLE 24 DESCRIPTIVE STATISTICS FOR GAIN SCORES ON THE EI TEST PER GROUP

TYPE OF ASSESSMENT TASK	*GROUP*	*N*	*Mean*	*SD*	*Median*	*IQR*
EI test (overall)	Information transmission	30	2.53	3.40	2.00	5.00
	Decision-making	30	2.00	3.07	1.00	4.00
EI test (grammatical)	Information transmission	30	1.96	2.34	1.00	4.00
	Decision-making	30	1.36	2.52	1.00	2.25
EI test (ungrammatical)	Information transmission	30	.56	1.94	.00	1.25
	Decision-making	30	.63	1.62	.00	1.00

Note: The total score for the EI test was 24 (12 for grammatical items, 12 for ungrammatical ones).

TABLE 25 GROUP DIFFERENCES IN GAIN SCORES

TYPE OF ASSESSMENT TASK	*Mann-Whitney U*	*Sig. p*	*Effect size r*
Oral production	311.50	.037	.26
Written production	278.00	.009	.33
EI (overall)	409.00	.542	.07
EI (grammatical)	383.50	.321	.12
EI (ungrammatical)	425.00	.693	.05

Research Question 3

The third research question asked the extent to which task complexity influenced the relationship between modified output produced in response to recasts and development in knowledge of the present third person singular. Table 26 depicts the relationship between the amount of modified output and L2 development in the simple condition (information transmission group) and Table 27 shows this relationship for the complex condition (decision-making group). In the simple condition, modified output positively correlated with development on the oral production test (rho =.443, p =.014) and on the written production test (rho =.444, p =.014). In the complex condition, there was a positive correlation between modified output and subsequent development only for the oral production test (rho = .450, p = .013). Interestingly, in both conditions, the correlations became strong in the second tasks of each group after the learners had received large numbers of intensive recasts (for the information transmission group rho =

.672, p = <.001 on the oral production test and rho = 544, p = .003 on the written production test; for the decision-making group rho =.456, p = .013 on the oral production test).

TABLE 26 CORRELATIONS BETWEEN MODIFIED OUTPUT AND L2 DEVELOPMENT (SIMPLE CONDITION)

Outcome measure	*N*	*Modified output during information transmission Task 1*		*N*	*Modified output during information transmission Task 2*		*N*	*Modified output during both tasks*	
		rho	p		rho	p		rho	p
Oral production test	30	.425	.019	27	.672	<.001	30	.443	.014
Written production test	30	.450	.013	27	.544	.003	30	.444	.014
EI test	30	.055	.775	27	.177	.378	30	.122	.522
EI grammatical	30	.186	.326	27	.138	.493	30	.204	.280
EI ungrammatical	30	-.236	.209	27	.092	.650	30	-.147	.440

TABLE 27 CORRELATIONS BETWEEN MODIFIED OUTPUT AND L2 DEVELOPMENT (COMPLEX CONDITION)

Outcome measure	*N*	*Modified output during decision-making Task 1*		*N*	*Modified output during decision-making Task 2*		*N*	*Modified output during both tasks*	
		rho	p		rho	p		rho	p
Oral production test	30	.369	.045	29	.456	.013	30	.450	.013
Written production test	30	.280	.134	29	.208	.278	30	.284	.128
EI test	30	.060	.751	29	.126	.514	30	.130	.492
EI grammatical	30	.106	.578	29	.226	.239	30	.178	.346
EI ungrammatical	30	-.210	.265	29	-.211	.271	30	-.205	.276

Research Question 4

The aim of the fourth research question was to illuminate whether aptitude was related to modified output produced after recasts, irrespective of task complexity. According to Table 28, there was no correlation between aptitude and the amount of modified output when the simple and complex conditions were examined as one group.

TABLE 28 CORRELATIONS BETWEEN APTITUDE AND MODIFIED OUTPUT FOR BOTH CONDITIONS

	N	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
		rho	p	rho	p	rho	p	rho	p
Modified output task 1	60	-.138	.292	-.190	.145	.220	.092	.132	.316
Modified output task 2	56	-.108	.428	-.077	.572	.188	.164	.041	.766
Modified output both tasks	60	-.181	.167	-.183	.162	.175	.181	.088	.503

Note: Fewer than 60 students modified their output in the second task as some of them produced the target feature accurately and no recasts were provided.

Research Question 5

The purpose of the fifth research question was to elucidate whether aptitude was related to the effectiveness of recasts in promoting knowledge of the target construction, regardless of task complexity. No correlation was found between aptitude and L2 development (Table 29).

TABLE 29 CORRELATIONS BETWEEN APTITUDE AND L2 DEVELOPMENT FOR BOTH CONDITIONS

Outcome measure *(N=60)*	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
	rho	p	rho	p	rho	p	rho	p
Oral production test	-.103	.432	.243	.061	.089	.498	-.030	.818
Written production test	-.132	.313	.167	.202	.138	.294	-.055	.677
EI test	-.116	.378	.075	.567	.217	.095	.058	.661
EI grammatical	-.176	.178	-.081	.540	.158	.229	.218	.095
EI ungrammatical	.123	.349	.185	.158	.225	.084	-.149	.256

Research Question 6

The sixth research question asked whether task complexity influenced the relationship between L2 aptitude and modified output produced after recasts. Table 30 demonstrates the results for the simple condition and Table 31 for the complex condition. Apart from a very weak positive correlation between phonetic coding and modified output in the complex condition (rho = .363 and p = .049), no other positive correlations were found between the two constructs for either the simple or the complex tasks. Nonetheless, it should be noted that there was a stronger positive correlation between phonetic coding and modified output in the first task in the complex condition (rho = .422, p = .020). The Tables below also show that, in both conditions, learners who modified their output were fewer than 30 in the second task as no recasts were needed for some participants due to their targetlike production of the present third person singular (N = 27 in the information transmission group, N = 29 in the decision-making group).

TABLE 30 CORRELATIONS BETWEEN L2 APTITUDE AND MODIFIED OUTPUT (SIMPLE CONDITION)

INFORMATION TRANSMISSION	*N*	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
		rho	p	rho	p	rho	p	rho	p
Modified output task 1	30	.013	.947	-.390	.033	-.006	.975	-.029	.880
Modified output task 2	27	.027	.894	-.095	.636	.056	.783	-.230	.249
Modified output both tasks	30	.004	.982	-.344	.062	-.021	.910	-.101	.595

TABLE 31 CORRELATIONS BETWEEN L2 APTITUDE AND MODIFIED OUTPUT (COMPLEX CONDITION)

DECISION-MAKING	*N*	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
		rho	p	rho	p	rho	p	rho	p
Modified output task 1	30	-.312	.093	-.093	.626	.422	.020	.309	.097
Modified output task 2	29	-.247	.197	-.155	.422	.293	.123	.321	.089
Modified output both tasks	30	-.352	.056	-.121	.525	.363	.049	.287	.124

Research Question 7

The seventh research question addressed whether task complexity influenced the relationship between L2 aptitude and development in knowledge of the present third person singular. No correlation was found between learners' development and L2 aptitude under the simple condition (Table 32). Meanwhile in the complex condition (Table 33), there was a strong positive correlation between the LLAMA D construct (i.e. ability to recognize sounds) and L2 gains demonstrated on the oral production test (rho = .539, $p = .002$). There was also a positive correlation between the LLAMA E construct (i.e. ability to associate sounds with symbols) and the learners' development on the written production test (rho = .455, $p = .012$). Finally, a positive correlation was found between the LLAMA E construct and development achieved on the EI test when composed of both grammatical and ungrammatical items (rho = .449, $p = .013$). There were also weak positive correlations between (1) LLAMA D scores and the learners' improvement on the written production test (rho = .386, $p = .035$) and (2) the LLAMA F component (i.e. grammatical inference) with development measured by the EI grammatical test (rho = .366, $p = .047$).

TABLE 32 CORRELATIONS BETWEEN L2 DEVELOPMENT AND L2 APTITUDE (SIMPLE CONDITION)

Outcome measure (N=30)	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
	rho	p	rho	p	rho	p	rho	p
Oral production test	-.017	.929	-.176	.353	-.090	.635	-.076	.689
Written production test	-.047	.804	-.166	.380	-.133	.485	-.144	.448
EI test	.170	.368	.157	.406	-.112	.554	-.107	.575
EI grammatical	.082	.668	.030	.875	-.093	.624	.067	.726
EI ungrammatical	.255	.173	.280	.134	.101	.594	-.184	.330

TABLE 33 CORRELATIONS BETWEEN L2 DEVELOPMENT AND L2 APTITUDE (COMPLEX CONDITION)

Outcome measure (N=30)	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
	rho	p	rho	p	rho	p	rho	p
Oral production test	-.194	.303	.539	.002	.291	.118	-.019	.923
Written production test	-.224	.234	.386	.035	.455	.012	-.006	.973
EI test	-.428	.018	.017	.928	.449	.013	.229	.224
EI grammatical	-.485	.007	-.113	.553	.328	.077	.366	.047
EI ungrammatical	-.034	.856	.113	.553	.323	.082	-.102	.591

Summary of Results (Study 1)

In summary, although the two groups did not differ in the amounts of modified output they produced, the simple condition, i.e. the information transmission group, achieved greater L2 gains on all tests in comparison to the complex condition, i.e. the decision-making group. Nonetheless, the differences between the two groups regarding their improvement were significant only on the oral and written production tests. Output modification positively correlated with subsequent L2 development gauged by the oral production test in both conditions, and by the written production test in the simple condition. Finally, in the simple condition, L2 aptitude was not associated with participants' improvement, while in the complex condition, several aptitude constructs positively correlated with learning benefits.

3.4.13 Discussion

Research Question 1

The first research question explored the potential influence of task complexity on the amount of modified output produced by the participants after receiving recasts. The hypothesis that simple tasks would lead to more modified output than complex tasks was not confirmed. In particular, in the simple condition, the students modified their output half of the time (51.45%) when they were provided with a recast, whereas in the complex condition, they modified their output less than half of the time (40.50%).

However, this difference was not significant. Thus, similar to the study conducted by Révész et al. (2011), task complexity did not influence the amount of modified output produced. Regarding the current empirical work, the rate of modified output was relatively low in both groups (i.e. simple and complex conditions). This might indicate that factors other than task complexity are involved.

First, as noted earlier, SLA researchers agree that output modification is a voluntary interactional move, and it cannot be considered a robust indicator of cognitive processes such as noticing. That is, learners might notice the linguistic target of feedback without correcting their errors (Ellis et al., 2001; Loewen, 2004). For example, remaining silent after a recast may implicate deeper processing of a target linguistic feature than repeating it immediately after the provision of feedback.

Second, the relatively low amount of modified output regardless of the cognitive demands of the tasks could be attributed to the interrogative mode of the recasts. Sheen (2006) showed that interrogative recasts were followed by a greater amount of uptake than declarative ones; however, recasts delivered in the interrogative mode induced a lower amount of modified output in comparison to declarative ones. In other words, the participants responded to interrogative recasts but their responses did not necessarily involve corrections of their initial errors. The interrogative recasts used in the present study often encouraged the learners to acknowledge their interlocutors' input (e.g. by replying "yes"); however, these responses are vague in terms of indicating noticing. Put differently, it is not clear whether: (1) the participants perceived the recasts as morphological CF targeting the present third person singular, and they engaged in cognitive comparisons between their erroneous output and the targetlike model supplied by the recasts or (2) they perceived recasts as confirmation checks intending to ask about the semantic content of their previous utterances. The former entails successful recasts whose corrective intention is unambiguous and, consequently, they enable the learners to attend to their linguistic target. In contrast, perceiving recasts as confirmation checks of meaning may deter learners from allocating attention to target linguistic errors. According to previous literature, recasts promote L2 development when learners interpret them as corrective moves focusing on form-meaning mappings rather than only meaning (Nicholas, Lightbown, & Spada, 2001).

Considering that modified output is a weak measure of noticing, the current study delved more deeply into the effectiveness of recasts and task complexity by formulating a second research question that focused on gain scores from a pretest to a posttest. The second research question is discussed below.

Research Question 2

The second research question investigated the combined effects of task complexity and recasts in promoting knowledge of the present third person singular. The hypothesis formulated for this question was confirmed; the learners benefited more from recasts delivered in the simple condition involving mere information transmission in comparison to the decision-making, complex condition that imposed greater reasoning demands. One reason why increased task complexity might not have promoted interlanguage development has to do with the nature of the target construction of the study. In particular, as explained earlier (section 3.4.3), the present third person singular is a non-salient and communicatively redundant feature, and hence it was not relevant to the task demands of the complex condition. In other words, the learners could cope with the greater communicative demands of the decision-making tasks without attending to the redundant allomorphs of the present third person singular addressed by recasts. Thus, it could be argued that increased task complexity may be more likely to assist development in the knowledge of linguistic features if those are relevant to task demands, such as question formation in Kim (2012).

A model that could explain the findings of the present study is Skehan's Limited Capacity Model (2009, 2014b). Following Levelt's (1989) model of speech production, Skehan contends that tasks with greater cognitive demands consume more attentional resources in the conceptualizer stage, whereby learners plan the content of their message and, consequently, less attention can be devoted to linguistic elements during formulation (see section 2.3.1). In contrast, simple tasks ease the cognitive load during performance, resulting in the allocation of more attention to linguistic features. The Limited Capacity Model could also have implications for the processing of feedback delivered during simple and complex tasks; when the cognitive burden of a task increases, less attention may be available to devote to the linguistic target of feedback, whereas tasks with lower cognitive demands may enable greater processing of the target construction. In the current study, the condition imposing lower cognitive demands (i.e. information transmission tasks) enabled the students to make greater use of their internal resources (i.e. attention, memory) and to engage in deeper processing of the present third person singular. Conversely, during complex decision-making tasks that involved greater reasoning demands, the learners might have directed their attention towards content related to decisions they had to make (match items with owners) rather than towards the target linguistic feature. In other words, it seems that while making decisions, the participants devoted greater attention to the communicative demands of the task that required them to choose a suitable item for each character and justify their decisions at the expense of processing the present third person singular, which is a communicatively redundant element not needed for successfully completing the task. Following Levelt's model, the decision-making tasks increased the cognitive load in the conceptualizer stage, leading to fewer attentional resources being available for processing of the linguistic target of the recasts. Hence, recasts delivered during cognitively complex tasks were less effective than those provided during simple tasks in promoting development of the present third person singular, which is a redundant feature.

Another possible explanation why recasts during complex, decision-making tasks were less successful than those supplied during simple, information transmission tasks may be related to the sources of input involved in each condition. Specifically, during decision-making tasks, the participants had to process more sources of input: both pictures depicting habits and several objects that belonged to the characters of the tasks. Conversely, no items were employed in the information transmission tasks and so the learners only had to process the content of pictures showing the characters' weekend activities. Révész (2009) and Révész et al. (2014) point out that tasks requiring the processing of several sources of input might be less beneficial in terms of processing CF as they might divert students' attention away from target linguistic areas.

An important parameter that should also be taken into account when attempting to shed light on the reasons why greater cognitive task demands impeded deeper processing of the target construction might be related to the learners' prior knowledge. As the pre-test scores revealed, the learners had no or very limited previous knowledge of the target construction prior to the treatment. In particular, the participants' mean scores were 3.99% and 3.75% on the oral production pretest, and 4.10% and 4.93% on the written production pretest for the information transmission and decision-making groups, respectively. If the participants had already partially mastered the present third person singular before the provision of recasts, they would probably have benefited from either the simple or complex tasks, and hence no significant differences would have been found in the gain scores of the two groups. The rationale for this prediction is that if the participants had achieved higher scores on the pretests, their prior knowledge might have relieved the processing demands posed under the complex condition, enabling them to devote greater attention to the recasts. Hence, it could be tentatively argued that learners in the first stages of acquiring a linguistic element benefit more from tasks with lower cognitive demands, especially when the target feature is redundant and/or is not easily noticed due to lack of salience.

It should also be underlined that the participants in both groups improved more on the written production test, followed by the oral production test, and they demonstrated very limited development on the EI test. Interestingly, although the recasts were supplied in the oral mode, the greatest development was attested to on the written test. At first glance, this finding seems surprising. According to skill acquisition theory, knowledge acquired from one skill is not easily transferred to another (e.g. from speaking to writing) (DeKeyser, 2007). In line with this argument, Révész (2012) found that the benefits of oral recasts were demonstrated to a greater extent on an oral test in comparison to written tests. In contrast, although the participants of the current thesis received oral recasts, they achieved their highest gain scores on a written test. Nonetheless, it could be argued that the students' improved performance in the written mode is still in line with skill acquisition theory. In particular, the L2 development of the participants shown on the written production test may not be a result of transferring their knowledge from one mode to another, but it may be attributed to the development of declarative in addition to procedural knowledge not yet automatized. Considering that the participants were in the first developmental stages of acquiring the target feature, as revealed by their pre-test performance, their L2 gains after the treatment were mainly reflected on the written production test, because it involved production under no time pressure and the possibility of output revision. It seems that those participants who noticed the target feature of the recasts during the treatment created initial representations of it in their long-term memory; however, in that early stage of development, they needed a sufficient amount of time to retrieve the linguistic target. Hence, the benefits of the oral treatment were exhibited when producing the present third person singular at their own pace on the written production test. Conversely, as the students' emerging knowledge of the target construction had not yet been automatized, they achieved less improvement on the oral production test, which involved greater time pressure as it was delivered in the oral mode, and very limited L2 gains on the EI test, which required oral production in only 10 seconds and the processing of predetermined input.

Research Question 3

The third research question examined the possible influence of task complexity on the relationship between modified output and development in knowledge of the present third person singular. Considering that previous studies have shown mixed findings with respect to whether modified output is an indicator of subsequent development (Loewen & Philp, 2006; Mackey & Philp, 1998 vs Egi, 2010; Loewen, 2005), no hypothesis was formulated with respect to the third research question. Although Révész et al. (2011) demonstrated that task complexity influences the extent to which modified output predicts L2 outcomes, a comparison between the two studies may be misleading as different types of task manipulations were examined. In the present study, a positive correlation was found between modified output and L2 development measured by the oral production test in both conditions. Modified output also correlated positively with learners' improvement on the written production test only in the simple condition. No correlation was demonstrated between EI gain scores and learners' modified output in either the simple or the complex condition.

The relationship between modified output and development on the oral production test in both conditions may indicate that, irrespective of task complexity, when the participants responded to recasts by correcting their errors, they did not simply "parrot" the target model, rather they processed the linguistic target and, as a consequence, they managed to generalize the pattern in novel verbs and improve their scores on the oral production test. As modified output in both conditions was generated during oral interaction, it is not surprising that it was also associated with a test administered in oral mode.

Modified output also positively correlated with L2 gains exhibited on the written production test but only in the simple, information transmission condition. It could be argued that modified output during simple tasks entailed the learners processing the target feature at a deeper level, and hence they were more successful in internalizing it and in producing it in a different mode from the one used during interaction. In contrast, in the complex condition which required greater mental effort, the participants did not process the target allomorphs to a degree that would enable them to produce them in written mode. Considering the greater demands of the decision-making tasks, it is possible that while modifying their output, the decision-making group devoted less attention to the target linguistic feature, as they were preoccupied with the content of their utterances (i.e. the decisions they had made). However, this is only a presumption as the learners' perceptions were not captured via stimulated recalls or immediate reports.

Another possible explanation of why there was no correlation between modified output and L2 outcomes on the written production test in the complex condition could be related to the learners' aptitude. As the decision-making group's improvement was associated with their aptitude, measured by LLAMA D and E, it is possible that the learners noticed the target feature when producing modified output during the treatment; however, low aptitude may have hindered its production on the written test for two reasons. First, the written production test was the last outcome measure administered, and hence those students who modified their output had to retain the oral allomorphs longer. That is, low aptitude related to the students' ability to recognize and retain oral patterns in LTM, as measured by LLAMA D, may not have enabled the learners to produce the present third person singular on the written production test as the features they may have noticed faded away. Second, the written production test involved sound-symbol connections for the present third person singular. Low aptitude in associating sounds with symbols, as assessed by LLAMA E, may not have allowed learners who noticed oral allomorphs to produce them in written mode. However, these interpretations should be made cautiously as modified output does not necessarily entail noticing linguistic elements (see section 2.1.4).

Research Questions 4 and 5

The fourth and the fifth research questions explored whether there was a relationship between (1) aptitude and modified output and (2) aptitude and L2 development when the information transmission and decision-making conditions were examined as one group (N=60). The hypothesis was that learners with higher aptitude would benefit more from recasts compared to low aptitude learners. Nonetheless, the study demonstrated that the participants' aptitude correlated with neither modified output nor their gain scores. In other words, higher aptitude did not facilitate the production of more modified output or greater L2 development when task demands were not taken into account. However, as explained earlier, previous studies have demonstrated that aptitude is associated with L2 outcomes only in certain learning conditions that involve explicit or implicit feedback (Li, 2013, 2015; Sheen, 2007; Trofimovich et al. 2007; Yilmaz, 2013; Yilmaz & Granena, 2015) and explicit or implicit instruction (DeGraaff, 1997; Erlam, 2005; Hwu & Sun, 2012; Robinson, 1997; VanPatten et al., 2013). Another factor that may also influence potential correlations between aptitude and L2 benefits is whether CF is supplied during simple or complex tasks. Hence, in the current thesis, it was highly important to delve deeper into the role of aptitude under different experimental conditions. To this end, two additional research questions were formulated intending to elucidate whether task complexity influences the relationship of aptitude with either modified output or learners' improvement. These research questions are discussed below.

Research Question 6

The aim of the sixth research question was to illuminate whether task complexity influenced the relationship of aptitude with modified output. The hypothesis that high aptitude students would produce more modified output than low aptitude students, especially under complex task conditions, was not confirmed. No correlation was found between aptitude and modified output in either the simple or the complex condition. This finding may have three possible explanations.

First, learners may manage to modify their output regardless of L2 aptitude after receiving explicit recasts. Drawing on previous literature, the recasts employed in the current study were explicit as they were partial and only involved one change. That is, they were not a full rephrasing of the learners' utterances but focused on single errors (see Sheen, 2006 for a review). Consequently, due to their explicitness, it seems that even learners with low aptitude managed to decode the target linguistic problem and modify their initial errors.

Second, recasts were delivered in a persistent manner during focused tasks which elicited frequent production of the target feature. Hence, both groups had great exposure to the present third person singular during the treatment. The large number of CF episodes addressing the target construction may have resulted in eliminating the influence of aptitude and, consequently, even learners with low aptitude modified their erroneous utterances. This presumption is reinforced by the fact that, in the complex condition, the LLAMA E construct positively correlated with the amount of modified output produced during the first treatment task; however, this correlation was not found in the second treatment task after the learners received many recasts (see table 31).

Finally, another possible reason why modified output did not correlate with L2 aptitude is that the former cannot serve as a reliable measure of noticing (Ellis et al., 2001). Considering the low amount of modified output produced by both groups, it could be argued that learners with high aptitude may have noticed the feedback and have processed the target feature without modifying their non-target-like utterances. Hence, examining the relationship between aptitude and gain scores provides further interesting insights into the role of aptitude in L2 outcomes. To this end, a seventh research question was formulated. Its results are discussed below.

Research Question 7

The purpose of the seventh research question was to explore whether task complexity influences the relationship between L2 aptitude and development in knowledge of the present third person singular. The hypothesis that L2 aptitude would be associated with L2 development in the complex condition was confirmed. In particular, while aptitude positively correlated with learners' L2 gains in the complex condition, no such relationship was demonstrated in the simple condition. Nonetheless, it should be noted that great variability was found in the gain scores of the information transmission group on both the oral and written production tests. In particular, the interquartile ranges were 57.19 for the oral production test, and 100.00 for the written production test (see Table 23). This indicates that some students exhibited dramatic improvements (e.g. from 0% on the pretest to 100% on the posttest of the written production measure), while others did not benefit from the recasts at all (e.g. they maintained their 0% scores from the written production pretest to the posttest). This finding suggests that other factors not explored by the current research project may have impacted on gain scores in the simple condition, such as learners' motivation or other aptitude constructs, e.g. attention control and working memory capacity. Regarding the aptitude constructs measured by the LLAMA test, it seems that when receiving recasts during cognitively simple tasks, learners benefit from feedback regardless of their aptitude scores.

With respect to complex decision-making tasks, the relationship found between the participants' improvement from pretest to posttest and their aptitude may be attributed to several factors. First, drawing on the Limited Capacity Model, it could be argued that high aptitude can compensate for learning conditions that involve the processing of recasts during complex tasks which increase the cognitive load in the conceptualizer stage, leading to less attention being allocated to linguistic features addressed by recasts. In other words, although complex tasks impose great cognitive demands, possibly resulting in less attention to language and CF, learners with high aptitude can still benefit from feedback and exhibit improved performance. This finding also echoes Robinson's (2007a, 2011) argument that individual differences are expected to play a more important role when learners perform more complex tasks in comparison to simple ones.

Another explanation for the correlations between aptitude and L2 gains may be related to the target construction. As explained in section 3.4.3, the present third person singular is a communicatively redundant and non-salient feature not easily noticed by L2 learners. Nonetheless, having high aptitude may facilitate the acquisition of such features, especially when they are made more salient in the input (Skehan, 2014a). The explicit recasts in the current study placed allomorphs of the present third person singular at the end of utterances (e.g. “He makes?”); and consequently, they became more salient than they usually are during more natural interaction or under more implicit learning conditions (e.g. when implicit, full recasts are supplied). Hence, this type of intervention appeared to be beneficial for high aptitude learners.

Nonetheless, the question that arises is which aptitude components influenced L2 development in the complex condition. The two subtests of the LLAMA Aptitude Test that played a crucial role were LLAMA D and LLAMA E. A weaker correlation was also found between L2 gains and LLAMA F. Each of these components is discussed in greater detail below.

LLAMA D

A strong positive correlation was demonstrated between the improvement of the decision-making group on an oral production test and their performance on the LLAMA D test, which is a measure of learners' ability to recognize oral patterns (rho = .539, p = .002). This relationship could be attributed to the fact that both LLAMA D and the treatment tasks required the participants to process and remember oral input. For LLAMA D, the students were exposed to oral patterns that they had to retain in their LTM in order to recognize them later. Similarly, during treatment tasks, and while processing oral recasts, the learners had to decode oral allomorphs of the present third person singular, hold them in their LTM and produce them during an oral production test. An important difference between LLAMA D and oral recasts of the treatment is that the former involved only a recognition test of sounds the learners heard, whereas the oral production test required the participants to produce the target feature of the recasts. In other words, LLAMA D only entailed recognition of oral patterns, while the oral production test required the production of allomorphs. Despite this discrepancy, the high positive correlation between the LLAMA D scores and gain scores after receiving oral recasts indicates that the learners' ability to process and remember sounds as measured by the LLAMA D test, is associated with cognitive processes implicated when receiving oral recasts addressing sounds/ allomorphs such as /s/, /z/, / əz /. Put differently, the relationship between LLAMA D and learners' improvement reveals the facilitative role of the ability to recognize sounds in the development of morphological features. This finding is in line with Granena's (2013) argument that what LLAMA D measures may be related to the acquisition of morphology. It also provides support for Robinson's (2001a, 2002, 2007a, 2012) prognosis that the effectiveness of recasts in terms of leading to L2 benefits is related to learners' memory for contingent speech.

A weaker positive correlation was also demonstrated between the learners' LLAMA D scores and their improvement on a written production test (rho = .386, $p = .035$). This suggests that learners' ability to recognize oral patterns may facilitate the development of declarative in addition to procedural knowledge. In other words, once learners with higher LLAMA D scores noticed the sounds/ allomorphs, they were able to produce them under conditions that imposed no time pressure in written mode.

Nonetheless, not surprisingly, the correlation between LLAMA D and gain scores on a written production test was relatively weak. As explained in the discussion section for the second research question, knowledge gained from one mode is not easily transferred to another (e.g. from speaking to writing) (DeKeyser, 2007). Hence, aptitude in the LLAMA D construct appears to be more relevant to gains in oral mode.

It should also be emphasized that the participants in the present study had very limited prior knowledge of the target feature, and they received explicit recasts in a laboratory in an intensive manner. Hence, the ability to recognize oral patterns might play a pivotal role in processing oral recasts, at least when learners are in the first stages of acquiring a construction and when they are provided with explicit recasts. However, it is unclear whether the LLAMA D construct would contribute to the L2 gains of learners in more advanced stages or if more implicit feedback was employed (e.g. full recasts under more natural interaction). Moreover, the current study cannot draw conclusions about whether the LLAMA D component is implicated in different areas of grammar (e.g. constructions that involve structural complexity, such as question formation). Saito (2017) found no correlation between morphological accuracy during oral production and learners' LLAMA D scores. Hence, it is highly possible that sound recognition differentially influences the development of grammatical features (e.g. those involving redundant, non-salient sounds such as the present third person singular and those entailing complex rules and/or structural complexity).

LLAMA E

Positive correlations were found between the decision-making group's development and their performance on the LLAMA E test, which is a measure of the ability to associate sounds with their symbols (phonetic coding). This finding is in line with other studies that have utilized LLAMA E in order to explore its role in L2 outcomes. In particular, Saito (2017) also demonstrated a positive relationship between the ability to associate sounds with symbols and morphological accuracy. Moreover, similar to the present empirical work, Yilmaz and Koylu (2016) showed that phonetic coding was associated with the efficacy of feedback. Although Yilmaz and Koylu employed explicit feedback (i.e. the provision of a target-like construction preceded by the comment "You should say…"), the results of the two studies are comparable. First, the recasts in the current project were relatively explicit (see section 3.4.6); and second, both studies employed feedback that provided the participants with a target-like feature. Hence, in agreement with Yilmaz and Koylu, the present study also shows that higher phonetic coding assists learners to process target linguistic constructions modelled by oral feedback. Nonetheless, in Yilmaz and Koylu, a relationship between phonetic coding and feedback was demonstrated when L2 learning was measured by a timed oral production test, whereas the present study uncovered this relationship only when L2 development was gauged by an EI and a written production test, but not an oral production test.

Regarding the positive correlation between phonetic coding and L2 improvement exhibited on a written production test (rho = .455, $p = .012$), it provides an insight into the type of knowledge the learners of the present empirical work developed. Saito (2017) has argued that "phonemic coding ability allows L2 learners to deconstruct words into phonetic units and analyse the form (pronunciation, morphology) and meaning aspects of words separately" (p. 683). In line with Saito's argument, it seems that participants who had higher phonetic coding in the current thesis engaged in such analysis and managed to develop declarative knowledge when receiving oral recasts targeting morphological features. These gains in their declarative knowledge were reflected in their improved performance on a written production test. Unlike the oral production test, the written test allowed (1) production of the target feature without time pressure and (2) subsequent revisions of the learners' output. Hence, the participants had more time at their disposal to access the declarative knowledge they developed during the treatment after receiving oral recasts. Considering that the LLAMA E subtest involves a study phase that allows learners to analyse language in a more explicit manner (Granena, 2013; Saito, 2017), its relationship with L2 gains on a test that served as a measure of both declarative and procedural knowledge was not a surprise. In Yilmaz and Koylu's (2016) study, although a relationship was found between learners' phonetic coding and performance on a timed, oral production test, learning was assisted by explicit feedback and it was captured by a test that involved oral production under relatively controlled conditions, rather than more spontaneous L2 production.

Interestingly, the LLAMA E component also correlated positively with the gains of the decision-making group on an EI test composed of both grammatical and ungrammatical items (rho = .449, p = .013)[6]. This finding supports Suzuki and DeKeyser's (2015) argument that EI tests may have elicited, at least in part, the use of automatized explicit knowledge. Although the EI test was employed as a measure of procedural knowledge in the process of automatization, the use of declarative knowledge while taking the test cannot be excluded. As the students in the current study were from an EFL environment that has “trained” them to analyse language explicitly, some of them may have accessed their automatized explicit knowledge during the 10 seconds they had to produce the utterances of the test. With regard to the role of aptitude, the positive correlation between the participants' improved performance on the EI test and LLAMA E demonstrates that the participants who managed to resort to their explicit knowledge were students with high phonetic coding ability. The argument that the students deployed their automatized explicit knowledge while taking the EI test is also reinforced by the positive correlation between learners' LLAMA F scores measuring grammatical inferencing with L2 gains demonstrated in the grammatical sentences of the EI test (see the section below).

[6] When gain scores were examined separately for grammatical and ungrammatical items on the EI test, they did not correlate with the learners’ LLAMA E scores.

Interestingly, the facilitative role of the LLAMA E construct in the complex condition was also revealed during the treatment while receiving recasts. In particular, apart from the positive correlations found between LLAMA E and the gain scores on the written production and EI tests, there was also a negative correlation between the participants' phonetic coding and the number of recasts supplied during the treatment tasks (rho = -.420, p = .021). That is, those learners who achieved higher scores on the LLAMA E test needed fewer recasts during the treatment. In other words, higher phonetic coding assisted learners to notice the present third person singular, and hence to produce the target feature accurately in novel verbs without the support of feedback.

The important role of the LLAMA E construct in the current study is also in line with Skehan's (2002) suggestion that phonetic coding is expected to influence L2 outcomes in the first stages of learners' development. The participants in the present empirical work were young learners with limited prior knowledge of the target construction and with low levels of proficiency. Hence, higher phonetic coding seems to assist young learners in the first developmental stages to process oral recasts targeting morphology, and to achieve greater L2 gains in comparison to their low aptitude counterparts.

LLAMA F

A weak positive correlation was found between LLAMA F, which is a measure of grammatical inferencing, and the learners’ development on the grammatical items of the EI test (rho = .366, p = .047). In other words, language analytic ability only facilitated development in those items that provided the target feature (grammatical sentences), while it did not affect gain scores in those items that omitted the linguistic target so that the learners had to produce it by themselves (ungrammatical sentences). Similar to phonetic coding (LLAMA E), language analytic ability captured by LLAMA F is also linked to explicit learning (Granena, 2013; Saito, 2017). Hence, the positive correlation between LLAMA F and gain scores on the grammatical items in the EI test echoes Suzuki and DeKeyser's (2015) argument that EI tests may, at least partly, gauge automatized explicit knowledge. Nonetheless, surprisingly, greater analytic ability did not assist in the correction of ungrammatical items. This finding may be attributed to the participants’ stage of acquiring the target construction. Considering that the students had very low pretest scores, even learners with high language analytic ability were not in a developmental stage that enabled them to correct ungrammatical items provided in oral mode and under time pressure.

However, it should be underlined that the positive relationship between learners' analytic ability and L2 gains in grammatical items cannot be generalised to different outcome measures administered under different conditions. Unlike the result of the current study showing that learners' analytic ability, which entails the use of explicit cognitive mechanisms, assists L2 benefits on grammatical items in a timed EI test, Ellis (2005a) demonstrated that only ungrammatical items were a measure of learners' explicit knowledge gauged by an untimed GJT delivered in written mode. This indicates that under conditions with no time pressure (e.g. untimed, written GJT), learners may be able to employ their explicit knowledge and correct ungrammatical items. In contrast, when test-takers have to correct ungrammatical, oral stimuli under time pressure, they may not be able to utilize their strong analytic ability and improve their scores, especially in the first stages of L2 acquisition. Hence, testing conditions (e.g. timed, oral EI test in the present study vs untimed, written GJT in Ellis's study) may influence the degree to which L2 learners deploy their explicit knowledge.

Interestingly, apart from the weak correlation found between LLAMA F and the grammatical items in the EI test, no other correlations were demonstrated. In other words, the learners' development on the oral and written production tests was not associated with their language analytic ability. This finding was unexpected, as the target construction was a grammatical element, though it could be attributed to several factors.

First, a possible interpretation could be associated with the type of linguistic target. Previous studies exploring the relationship between aptitude and CF have shown that the extent to which learners with high language analytic ability benefit from recasts is related to the target linguistic feature. In particular, as explained earlier (section 2.4.7), it has been demonstrated that even learners equipped with strong language analytic ability do not benefit from recasts when non-salient or complex linguistic constructions are addressed (Sheen, 2007; Yilmaz, 2013; Yilmaz & Granena, 2015). In contrast, recasts have been found to be more effective for learners with high language analytic ability when the linguistic target refers to simple constructions with clear form-meaning mappings (Li, 2013; Yilmaz, 2013).

In the current study, the target construction is the present third person singular, which requires the connection of three allomorphs with the meaning of a single morpheme that is communicatively redundant. Considering that previous research on recasts has revealed that high language analytic ability is associated with the development of constructions with clear form-meaning connections (Li, 2013; Yilmaz, 2013), it could be argued that when these connections should be made for three different allomorphs that are also communicatively redundant, learners may not resort to their language analytic ability as they can convey the meaning of their utterances successfully without the redundant features.

Nonetheless, apart from the difficulty regarding form-meaning connections, the rule of the present third person singular involves a relatively simple pattern: the allomorphs /s/, /z/, and /əz/ are added to the base form of the verb. In other words, it is not a feature that involves structural complexity like some other constructions (e.g. comparatives). If we classify the present third person singular as what Ellis (2006) called "features for which ready rules-of-thumb are available" (p. 458), (e.g. past tense -ed), high grammatical inferencing ability might not play a crucial role in figuring out the linguistic pattern. Similarly, Saito (2017) predicted that analytic ability measured by LLAMA F is expected to facilitate "more diverse, sophisticated, and complex lexicogrammar usage beyond the use of frequent words and simple grammatical structures" (p. 670).

In light of the above arguments, the development of grammatical features such as the present third person singular that involve simple rules in terms of structural complexity may not be associated with learners' having strong analytic ability. On the contrary, when these features contain communicatively redundant and non-salient sounds, L2 gains may be linked to other aptitude constructs, such as recognizing sounds (LLAMA D) and phonetic coding (LLAMA E). Hence, in the current study, learners with low analytic ability but with high aptitude in constructs related to the processing of sounds (LLAMA D and E) were those who exhibited development in the present third person singular allomorphs after receiving recasts, whereas learners' high language analytic ability could probably not compensate when low scores on LLAMA D and E were achieved.

Another possible explanation why LLAMA F scores were not associated with L2 development may be related to the type of feedback delivered. During the treatment, only recasts that modelled the target-like feature were supplied (e.g. He makes?). This type of feedback may have not been sufficient for the learners to activate their language analytic ability. In contrast, language analytic ability may assist development when more explicit feedback is employed (e.g. directly rejecting learners' erroneous utterances or providing metalinguistic information). For example, Sheen (2007) and Yilmaz and Granena (2015) found that language analytic ability was not associated with development in the recast condition, whereas both of these studies revealed a positive relationship between language analytic ability and L2 gains under the condition involving explicit feedback (i.e. metalinguistic feedback in Sheen (2007) and explicit correction in Yilmaz & Granena (2015)).

The findings of study 1 discussed above have implications that are useful at both a theoretical and a pedagogical level. These conclusions are presented in chapter 5. However, in CF and aptitude research, apart from task complexity, which was the focus of study 1, another variable that may also play an important role in the efficacy of recasts is the mode of interaction (i.e. oral recasts in FTF mode vs written recasts in SCMC mode). This variable was examined by a second study presented in the following chapter.

CHAPTER 4: STUDY 2

4.1 Aims and Rationale

The second study intends to explore the combined effects of mode of interaction and explicit recasts on modified output and L2 development in knowledge of the present third person singular, and whether these presumed links are associated with L2 aptitude. It also examines the relationship of modified output with L2 learning. Only a limited number of studies has delved into the effects of text-based CF delivered during online interaction on L2 outcomes (see section 2.2). Nonetheless, in all of these studies, the participants were adults. The novelty of the current experiment concerns four areas: First, the current thesis investigates the potential benefits of SCMC feedback for young EFL learners. Second, an under-researched linguistic construction is investigated, i.e. the present third person singular. As explained earlier, in section 2.2, written feedback provided during SCMC might be especially helpful for non-salient linguistic features not easily noticed during oral interaction. Nevertheless, more empirical studies are needed to show whether the SCMC environment is indeed effective in drawing learners' attention to such features. Third, although previous research has examined the relationship of modified output and L2 development in oral interaction, this relationship has not been examined in SCMC mode. Finally, the extent to which the mode of interaction influences potential relationships of aptitude to modified output and L2 gains has been the object of relatively limited research. Considering that different cognitive processes are implicated after receiving oral and written recasts in the FTF and SCMC modes respectively, it is possible that different aptitude constructs are involved in each learning environment.

4.2. Research Questions and Hypotheses

To address these areas, study 2 has three sets of research questions, all concerned with conditions where recasts remain constant: the first set asks the extent to which mode of interaction affects modified output and L2 development and their relationship. The second set focuses on how aptitude relates to modified output and L2 benefits. Finally, the third set looks into how the mode of interaction might influence the relationships of aptitude to modified output and L2 gains.

Mode of interaction, modified output and L2 development

1. To what extent does the mode of interaction affect the amount of modified output after recasts targeting the present third person singular?
2. What are the combined effects of mode of interaction and recasts on promoting knowledge of the present third person singular?
3. To what extent does the mode of interaction influence the relationship between modified output and development in knowledge of the present third person singular?

Aptitude, modified output, and L2 development

4. To what extent does aptitude relate to the amount of modified output after recasts targeting the present third person singular?
5. To what extent does aptitude relate to development in knowledge of the present third person singular, when recasts remain constant?

Mode of interaction, aptitude, modified output and L2 development

6. To what extent does the mode of interaction influence the relationship between L2 aptitude and modified output produced after recasts targeting the present third person singular?
7. To what extent does the mode of interaction influence the relationship between L2 aptitude and development in knowledge of the present third person singular, when recasts remain constant?

The hypothesis for the first and second research questions is that the SCMC mode will be more effective than the FTF mode in leading to modified output and L2 gains. The rationale for this prognosis is that, since SCMC feedback remains on screen, the salience of the target feature may increase. Moreover, interaction in the SCMC mode occurs at a lower pace than in FTF mode and it may afford greater processing time while relieving the cognitive burden imposed during oral communication (see section 2.2). Consequently, the SCMC group is expected to be more successful in noticing feedback and improving their test scores as compared to the FTF group. With respect to the third research question, similar to study 1, no hypothesis is made as previous research has indicated contradictory results about the relationship of modified output with L2 development. Furthermore, no study has examined whether the mode of interaction influences potential correlations between modified output and interlanguage development. Regarding research questions 4–7, it is presumed that the effects of aptitude may be lower in the SCMC condition as low aptitude students may benefit from SCMC text-based feedback due to the factors described above (i.e. salience, lower speed of interaction) (see sections 2.2 and 2.4 dedicated to interaction in the SCMC mode and individual differences, respectively).

4.3 Pilot study

With regard to the information transmission tasks performed in the oral mode, they were piloted on 10 learners of English from various L1 backgrounds and on 7 Greek students learning English as a second language in Greece (for more details see section 3.3). With respect to the information transmission tasks in SCMC mode, they were piloted on 5 Greek learners with similar profiles to those in the FTF group. The tasks were successful in eliciting the present third person singular in SCMC mode and generated a similar number of obligatory contexts of the target feature as those in FTF mode. An important difference between the two modes of interaction was that in FTF mode the length of the tasks was approximately 30 minutes, whereas for the SCMC group it was one hour. The students in the SCMC condition needed more time to plan their output and read the input (i.e. recasts). Due to the length of the SCMC tasks, the second study did not employ an EI test, which takes almost 25 minutes. The following section presents the design and methodology of the second study, the results and a discussion for each research question.

4.4 Research Design and Methodology of Study 2

4.4.1 Design

Similar to study 1, the second study also used a pretest-posttest design. After being administered a proficiency test, 60 participants were randomly assigned to one of two experimental groups (see Table 34). Both groups performed information-transmission tasks, but they differed as to whether they interacted in face-to-face (FTF) mode or engaged in synchronous computer-mediated communication (SCMC) with the researcher. In both conditions, the learners received interrogative recasts from the researcher in response to errors associated with the present third person singular. The learners' L2 development from pretest to posttest was assessed by two outcome measures that involved an oral production test and a written production test. As explained earlier, an EI test was not administered because the duration of the SCMC treatment was approximately one hour and the duration of the EI test was 25 minutes. The participants' L2 aptitude was measured by the LLAMA test (Meara, 2005).

TABLE 34 DESIGN OF STUDY 2

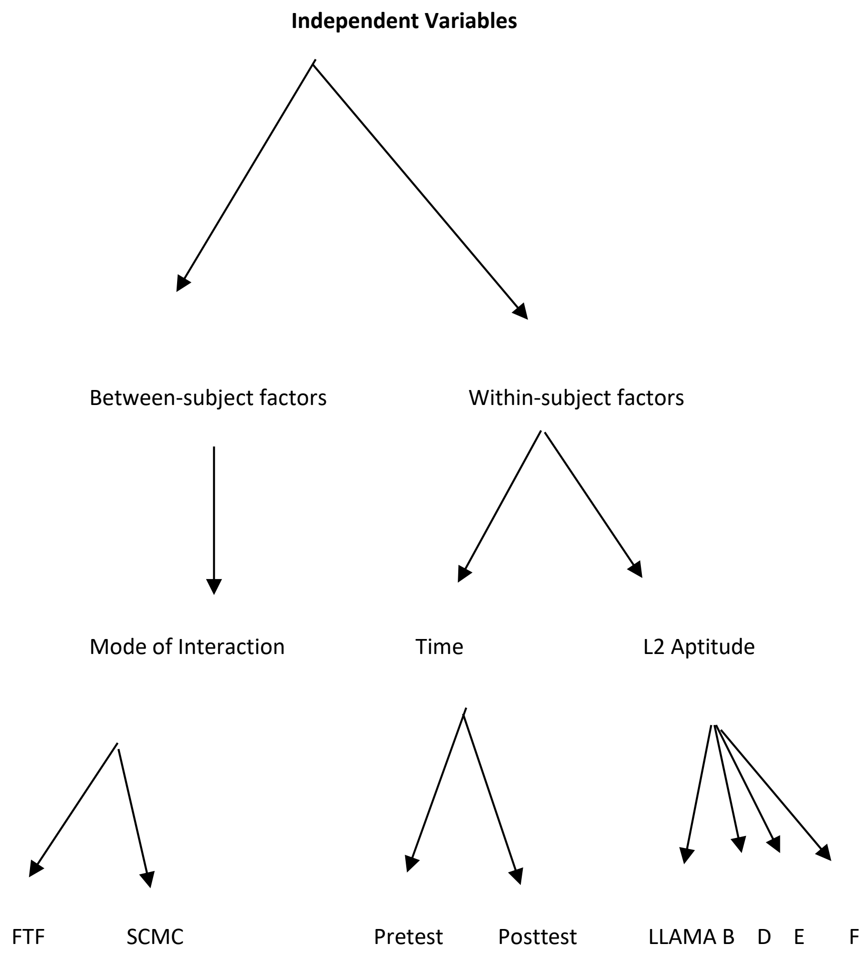
	FTF group N=30	SCMC group N=30
Proficiency test	ISE Foundation (A2) of Trinity College London.	
Pretest	Oral production	
	Written production	
Treatment 1	Task 1	
Treatment 2	Task 2	
Posttest	Oral production	
	Written production	
Aptitude	LLAMA B, D, E, F	

The first set of research questions (see section 4.2) was concerned with mode of interaction, modified output and L2 development. Specifically, for the first research question, the two groups (SCMC and FTF groups) were compared with regard to the amount of modified output they produced. The second research question was addressed by comparing the development of the two groups as measured by the assessment tools. As for the third research question, the relationship of L2 development and modified output was explored for the SCMC and FTF conditions.

The second set of questions were associated with aptitude, modified output and L2 development. In particular, for the fourth and fifth research questions, it was examined whether there were relationships between (1) aptitude and modified output and (2) aptitude and L2 development demonstrated by the participants, regardless of the mode of interaction. Finally, the third set of research questions concerned mode of interaction, aptitude, modified output and L2 development. Specifically, for the sixth and seventh research questions, potential correlations were investigated between (1) aptitude and modified output and (2) aptitude and L2 development. These correlations were calculated for each condition separately so as to illuminate the potential effects of the mode of interaction.

Overall, in the study there was (1) one between-subjects factor, mode of interaction, which consisted of two levels (face-to-face vs SCMC) (2) time as a within-subjects factor, which in turn involved two levels – a pretest and a posttest – and (3) two dependent variables – modified output and L2 development assessed by the two tests. The study also had an additional (4) within-subjects factor, L2 aptitude, which entailed four levels – LLAMA B, LLAMA, D, LLAMA E and LLAMA F (Meara, 2005). A schematic representation of the research design is provided in Figure 5.

Figure 5 Research Design of Study 2



Independent Variables:

-Mode of interaction

-Learners' L2 aptitude

-Time of the test (pre- and post-test)

Dependent Variables:

-Amount of modified output

-L2 development from pre- to the post-test composed of an oral production test and a written production test.

4.4.2 Participants

The participants were 60 Greek EFL learners. The FTF group consisted of the same participants as the information transmission group in the first study, apart from three students who were replaced in order to ensure that the pretest scores would be similar to those in the SCMC group. Overall, the profiles of the students participating in study 2 were similar to those in study 1. In particular, they attended EFL courses in a state school and several language schools in Greece, where the communicative approach was used. The selection criteria with regard to the participants' proficiency level and their prior knowledge of the target construction were the same as those for the participants in study 1. Specifically, the required proficiency level was elementary or pre-intermediate, as determined by a Trinity College London ISE Foundation (A2) test. Also, participants with limited prior knowledge were selected, i.e. scoring lower than 35% on any of the components of the pretest so as to avoid ceiling effects.

Background information for the participants in both groups (FTF and SCMC) is presented in Table 35 There were 34 females and 26 male learners, and their ages varied from 10.5 to 13 years (M=11.39, SD=.86). They were all native speakers of Greek; however, six of them were bilinguals, born in Greece but of Romanian (n=3), Albanian (n=2) or Russian (n=1) origin. Their length of learning English prior to the study ranged from 2 to 8 years (M=4.56, SD=1.19). The majority of the participants also reported learning a second language (i.e. German, French or Russian) (N=44). It should also be noted that none of the participants had ever lived in an English-speaking country before. In order to ensure that the learners in the SCMC condition were familiar with typing, they were asked how often they used keyboards and they all reported frequent use. A Mann-Whitney U test was run on the factors of age, length of learning English prior to the study and performance on the proficiency test. The two groups were comparable, as no significant differences were found for any of the three factors (Mann-Whitney U = 447.50, p = .969 for age; Mann-Whitney U = 436.50, p = .831 for length of learning English; Mann-Whitney U = 443.00, p = 916 for students' performance on a listening test). The effects sizes were r <.01 for age, r =.02 for length of previous English study and r =.01 for proficiency.

TABLE 35 DESCRIPTIVE STATISTICS FOR BACKGROUND INFORMATION PER GROUP

Group	*N*	*Gender*	*Age*			*Length of previous English study in years*			*Proficiency*		
			M (SD)	Median	IQR	M (SD)	Median	IQR	M (SD)	Median	IQR
FTF	30	16 M 14 F	11.38 (.92)	11.00	1.00	4.56 (1.22)	4.00	1.00	5.63 (2.18)	5.00	3.00
SCMC	30	10 M 20 F	11.40 (.82)	11.00	1.00	4.56 (1.17)	4.00	1.00	5.66 (2.07)	5.00	3.25

4.4.3 The Target Linguistic Construction

The target linguistic construction was the present third person singular. As explained earlier, this is a non-salient and communicatively redundant feature. Consequently, as a construction, it is not easily amenable to CF (see section 3.4.3).

4.4.4 The Treatment Task

The FTF group was the same as the one performing information transmission tasks in study 1 (apart from 3 participants, see section 4.4.2). The SCMC group was also asked to do the information transmission tasks used in study 1 (see section 3.4.4) but in an online environment. The reason why the information transmission tasks were selected for the second study was that these tasks were found to be more beneficial than the decision-making tasks in FTF mode and it was worth examining whether these benefits were also exhibited in SCMC mode.

4.4.5 Recasts

While carrying out information transmission tasks, the participants received interrogative, partial recasts from the researcher (e.g. "He watches?") when the target construction was produced inaccurately (see section 3.4.6). The FTF group received oral recasts as they engaged in oral interaction, whereas the SCMC group were provided with written recasts.

4.4.6 Outcome Measures

L2 development was measured by employing a pre-test and a post-test composed of the same oral and written production tests administered to the learners in study 1 (see section 3.4.7). However, in study 1, both groups interacted in oral mode, and hence a written test provided evidence about whether the potential benefits of oral recasts were apparent in written mode. In contrast, in the second study, one of the two experimental groups delivered recasts in the SCMC environment. Thus, the oral production test indicated whether potential learning benefits of written text-based recasts were transferred into oral mode.

4.4.7 Aptitude Test

As in Study 1, the learners were also asked to take a LLAMA test in order to measure their L2 aptitude (see section 3.4.9).

4.4.8 Data Collection

The process of data collection was exactly the same as in study 1 (section 3.4.10), except for the fact that, in study 2, the two experimental groups differed with regard to the mode of interaction (FTF vs SCMC) and an EI test was not administered due to the length of the tasks in the SCMC environment and constraints regarding participants' availability. The SCMC interaction was via Skype and it was recorded by screen capture software called SNAGIT. The experimental schedule and the duration of each stage are presented in Figure 6.

FIGURE 6 EXPERIMENTAL SCHEDULE OF STUDY 2

SESSION 1

Consent form and background questionnaire (5 minutes)

Proficiency test (10 minutes)

Pretests

Oral production (5 minutes)

Written production (15 minutes)

Aptitude test (LLAMA D, E and F)

(one-week interval)

SESSION 2

Treatment

FTF Group: Tasks 1 and 2 (30 min)

SCMC Group: Tasks 1 and 2 (1 hour)

(Break, approximately 5 minutes)

Aptitude LLAMA B

Posttests

Oral production (5 minutes)

Written production (15 minutes)

4.4.9 Data Analyses

Transcription

As in study 1, all the oral data (i.e. FTF tasks and tests) were transcribed by the researcher. The data from SCMC interaction were analysed using screen-capture software, i.e. SNAGIT. Ten per cent of the data were also transcribed by a second researcher to check inter-transcriber reliability. The data transcribed by the second researcher were selected through stratified random sampling across the experimental groups. A comparison of the two transcriptions focused on the target verbs. Discrepancies between the two transcriptions were related to (1) items differently transcribed and (2) items present in one transcription but absent from the other one. Inter-transcriber agreement was calculated by dividing the total number of items transcribed identically by the total number of items and it was high (.988). Cohen's kappa also revealed high inter-transcriber agreement (.962).

Coding and Scoring

The coding for modified output and the oral/written production tests was the same as in study 1 (see section 3.4.11). With regard to modified output, the learners were given an opportunity to produce uptake after receiving recasts in both groups. In SCMC mode, modified output was written on the screen, whereas in FTF mode it was oral.

Inter-Coder Agreement

Ten per cent of the data regarding modified output and test scoring were coded by a second researcher. As explained earlier, the research materials (tasks and tests) and the coding for output modification and test scores were the same for all experimental groups in study 1 and 2. Hence, the data selected to be coded twice represented all groups equally. A high level of agreement was found between the two coders (.98) and Cohen's kappa values were also high: .93 for modified output, .95 for the oral production test and .96 for the written production test.

4.4.10 Results

Preliminary Analyses

The present section describes the preliminary statistical analyses conducted before addressing the research questions. The aim of the analyses was to examine (1) whether the gain scores were normally distributed and (2) if the two groups were similar in their prior knowledge of the target construction as measured by pretests, in their L2 aptitude gauged by a LLAMA test and in the number of OCs of the target construction while interacting in the FTF and SCMC environments. These steps are discussed below.

Similar to study 1, before answering the research questions of study 2, a Shapiro-Wilk test was used to explore the normality of the distributions of the gain scores (Table 36, $p < .05$). All tests reached significance, thus non-parametric tests were required for the analyses. Similar to study 1, study 2 utilized a Mann-Whitney U test, and the effect sizes were computed by employing the formula $r = z/\sqrt{N}$. Drawing on Plonsky and Oswald (2014), the effect sizes were interpreted as small when *r* was close to .25, medium when *r* was close to .40 and large when *r* reached .60.

TABLE 36 TEST OF NORMALITY (SHAPIRO-WILK)

TYPE OF ASSESSMENT TASK	*GROUP*	*Shapiro-Wilk Sig.*
Oral production test	FTF	.002
	SCMC	<.001
Written production test	FTF	<.001
	SCMC	<.001

Furthermore, in order to ensure that the two experimental groups were comparable, before answering the research questions, a series of Mann-Whitney U tests were conducted to examine (1) whether the two groups were significantly different in their prior knowledge of the target feature, (2) whether the numbers of OCs of the target construction created in the FTF and SCMC modes were similar and (3) whether there were any significant differences between the two groups in their L2 aptitude.

First, it was explored whether the two groups were different in their knowledge of the target construction prior to the treatment. Table 37 presents descriptive statistics for the performance of the two groups on the pretest and Table 38 indicates that there was no significant difference between them with regard to their prior knowledge (Mann-Whitney U = 445.50, p = .924 for the oral production test; Mann-Whitney U = 414.50, p = .452 for the written production test). The effect sizes were also small (r = .01 for the oral production test, r = .09 for the written production test). Consequently, potential gains after the treatment could be attributed to the combined effects of mode of interaction and recasts.

TABLE 37 DESCRIPTIVE STATISTICS FOR THE TWO GROUPS' PRETEST SCORES

TYPE OF ASSESSMENT TASK	*GROUP*	*N*	*Mean (%)*	*SD*	*Median*	*IQR*
Oral production test	FTF	30	2.09	4.75	.00	.00
	SCMC	30	1.56	3.48	.00	.00
Written production test	FTF	30	3.25	7.43	.00	.1.92
	SCMC	30	2.15	5.69	.00	.00

TABLE 38 BETWEEN GROUP DIFFERENCES ON PRETEST SCORES

TYPE OF ASSESSMENT TASK	*Mann-Whitney U*	*Sig. p*	*Effect size r*
Oral production pretest	445.50	.924	.01
Written production pretest	414.50	.452	.09

Second, the researcher examined whether the two modes were similar regarding their numbers of OCs of the target feature so as to confirm that the two groups had equal opportunities to receive recasts. Table 39 shows descriptive statistics for the numbers of OCs created in the two conditions.

Table 40 demonstrates that the numbers of OCs between the two groups were not significantly different (Mann-Whitney U = 396.50, p = .419). The effect size was small (r = .10).

TABLE 39 DESCRIPTIVE STATISTICS FOR THE NUMBERS OF OCS PER GROUP

GROUP	*Number*	*Mean*	*SD*	*Median*	*IQR*
FTF (TASK 1)	448	14.93	2.39	15.00	1.25
FTF (TASK 2)	451	15.03	1.69	15.00	1.00
FTF BOTH TASKS	899	29.96	3.95	30.00	2.25
SCMC (TASK 1)	456	15.2	1.78	15.00	2.00
SCMC (TASK 2)	444	14.8	2.14	15.00	1.00
SCMC BOTH TASKS	900	30.00	3.67	30.00	3.00

TABLE 40 BETWEEN GROUP DIFFERENCES IN THE NUMBERS OF OCS

GROUP	*Number of OCs*		
	Mann-Whitney U	*Sig. p*	*Effect size r*
TASK 1	374.50	.252	.14
TASK 2	417.00	.599	.06
BOTH TASKS	396.50	.419	.10

Finally, the researcher investigated whether the two groups were comparable in terms of their L2 aptitude. Table 41 has the descriptive statistics for the learners' L2 aptitude per group, and Table 42 reveals that there were no significant differences between the two groups (Mann-Whitney U = 441.00, p = .893 for LLAMA B; Mann-Whitney U = 377.00, p = .275 for LLAMA D; Mann-Whitney U = 340.00, p = .101 for LLAMA E; Mann-Whitney U = 441.50, p = .898 for LLAMA F). The effect sizes were: r =.01 for LLAMA B, r =.14 for LLAMA D, r =.21 for LLAMA E, r =.01 for LLAMA F.

TABLE 41 DESCRIPTIVE STATISTICS FOR L2 APTITUDE SCORES PER GROUP

LLAMA COMPONENT	*GROUP*	*N*	*Mean*	*SD*	*Median*	*IQR*
LLAMA B	FTF	30	31.66	17.23	30.00	25.00
	SCMC	30	30.83	16.08	32.50	20.00
LLAMA D	FTF	30	21.83	11.40	20.00	15.00
	SCMC	30	18.50	11.82	17.50	20.00
LLAMA E	FTF	30	41.00	26.04	40.00	32.50
	SCMC	30	30.33	24.84	25.00	52.50
LLAMA F	FTF	30	21.00	20.73	20.00	40.00
	SCMC	30	23.00	24.37	20.00	35.00

TABLE 42 BETWEEN GROUP DIFFERENCES IN L2 APTITUDE

LLAMA COMPONENT	*Mann-Whitney U*	*Sig. p*	*Effect size r*
LLAMA B	441.00	.893	.01
LLAMA D	377.00	.275	.14
LLAMA E	340.00	.101	.21
LLAMA F	441.50	.898	.01

In summary, there were no significant differences between the two groups with respect to their prior knowledge of the target construction, OCs of the target feature in the FTF and SCMC conditions and participants' L2 aptitude. Hence, the two groups were comparable and potential gains could be attributed to the combined effects of recasts and mode of interaction.

After confirming that the two groups were comparable according to the above factors, a series of Mann-Whitney U tests were run in order to answer the research questions: (1) whether there were any significant differences between the two groups in the amount of modified output following recasts, (2) whether there were any significant differences between the two groups regarding their gain scores on each test (i.e. oral production and written production tests). Effect sizes were also calculated for each of these questions. Moreover, Spearman tests were run to explore whether there was a relationship between (1) the amount of modified output and subsequent L2 development per group, (2) aptitude and modified output irrespective of mode of interaction, (3) aptitude and L2 development regardless of mode of interaction, (4) aptitude and modified output for each group separately and (5) aptitude and L2 development for each group. The results for each research question are presented in the following section.

Research Question 1

The first research question asked the extent to which mode of interaction affected the amount of modified output produced by L2 learners after the provision of recasts targeting the present third person singular. Descriptive statistics for the amount of modified output per group are given in Table 43.

The FTF condition generated more modified output than the SCMC condition. In particular, the mean of output modification for the FTF group was 49.16% while for the SCMC group it was only 12.77%. As Table 44 demonstrates, the FTF group modified their output significantly more than the SCMC group after receiving recasts (Mann-Whitney U = 181.50, $p < .001$). The effect size was also relatively large (r= .52 for both tasks, r = .50 for task 1, r = .46 for task 2). In other words, oral recasts delivered during FTF interaction were more successful in inducing output modification than written recasts in the SCMC group. Moreover, regarding IQR, greater variability was found in the FTF group in compariso to the SCMC group (73.81 and 6.64, respectively).

TABLE 43 DESCRIPTIVE STATISTICS FOR MODIFIED OUTPUT PER GROUP

GROUP	*N*	*Mean (%)*	*SD*	*Median*	*IQR*
FTF (TASK 1)	30	48.48	37.32	45.45	66.67
FTF (TASK 2)	27	51.43	41.47	60.00	92.85
FTF BOTH TASKS	30	49.16	37.09	52.38	73.81
SCMC (TASK 1)	30	13.03	26.56	.00	7.55
SCMC (TASK 2)	28	12.78	30.96	.00	.00
SCMC BOTH TASKS	30	12.77	28.05	.00	6.64

TABLE 44 BETWEEN GROUP DIFFERENCES IN OUTPUT MODIFICATION

GROUP	*Mann-Whitney U*	*Sig. p*	*Effect size r*
TASK 1	192.50	<.001	.50
TASK 2	182.50	<.001	.46
BOTH TASKS	181.50	<.001	.52

Research Question 2

The second research question addressed the combined effects of mode of interaction and recasts on promoting knowledge of the present third person singular. Table 45 has descriptive statistics for the gain scores achieved per group. The FTF group outperformed the SCMC group on both tests. Specifically, the mean gain score for the FTF group was 27.35% on the oral production test and 49.24% on the written production test, whereas for the SCMC group they were 9.71% and 11.88% on the oral and written production tests, respectively. According to Table 46, the FTF group exhibited significantly greater gains than the SCMC group on both tests (Mann-Whitney U = 282.00, p = .010 on the oral production test; Mann-Whitney U = 256.50, p = .002 on the written production test). The effect size was medium (r = .33 for learners' L2 gains on the oral production test, and r = .39 for their L2 development on the written production test). In other words, the learners benefitted more from oral recasts during FTF interaction in comparison to written recasts in the SCMC environment. Interestingly, the benefits of oral recasts were transferred into written mode; however, this was not observed in the SCMC condition. Moreover, greater variability was observed in the FTF group as compared to the SCMC group. Regarding the gains seen on the oral production test, the IQR was 53.04 in the FTF group and 12.71 in the SCMC group. As for the gains on the written production test, the IQR was 100.00 in the FTF group and 3.43 in the SCMC group.

Table 45 Descriptive Statistics for Gain Scores per Group

Type of assessment task	*Group*	*N*	*Mean (%)*	*SD*	*Median*	*IQR*
Oral production test	FTF	30	27.35	27.50	25.38	53.04
	SCMC	30	9.71	19.32	.00	12.71
Written production test	FTF	30	49.24	46.51	58.33	100.00
	SCMC	30	11.88	26.89	.00	3.43

Table 46 Between Group Differences in Gain Scores

Type of assessment task	*Mann-Whitney U*	*Sig. p*	*Effect size r*
Oral production	282.00	.010	.33
Written production	256.50	.002	.39

Research Question 3

The third research question asked whether the mode of interaction influenced the relationship between modified output and development in knowledge of the present third person singular. Table 47 shows the relationship between modified output and L2 development in FTF mode, and Table 48 presents this relationship in SCMC mode. In the FTF environment, modified output positively correlated with learners' improvements on both oral and written production tests. In contrast, there was no relationship between output modification and L2 development in the SCMC condition. In other words, those learners who modified their output after oral recasts delivered during FTF interaction were more likely to exhibit L2 gains in comparison to learners who remained silent, replied "yes" or repeated the error in either the same or a different verb. Conversely, in the SCMC environment, modified output did not correlate with subsequent learning.

TABLE 47 CORRELATIONS BETWEEN MODIFIED OUTPUT AND L2 DEVELOPMENT (FTF MODE)

Outcome measure	*N*	*FTF Task 1*		*N*	*FTF Task 2*		*N*	*FTF BOTH TASKS*	
		rho	p		rho	p		rho	p
Oral production test	30	.416	.022	27	.675	<.001	30	.454	.012
Written production test	30	.402	.028	27	.563	.002	30	.419	.021

TABLE 48 CORRELATIONS BETWEEN MODIFIED OUTPUT AND L2 DEVELOPMENT (SCMC MODE)

Outcome measure	*N*	*SCMC Task 1*		*N*	*SCMC Task 2*		*N*	*SCMC BOTH TASKS*	
		rho	p		rho	p		rho	p
Oral production test	30	.093	.625	28	.181	.357	30	.092	.630
Written production test	30	.265	.158	28	.264	.175	30	.264	.159

Research Question 4

The fourth research question aimed to elucidate whether aptitude was associated with modified output produced after recasts, regardless of mode of interaction. According to Table 49, there was no relationship between aptitude and modified output when the FTF and SCMC conditions were explored as one group.

TABLE 49 CORRELATIONS BETWEEN APTITUDE AND MODIFIED OUTPUT FOR BOTH CONDITIONS

	N	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
		rho	p	rho	p	rho	p	rho	p
Modified output task 1	60	-.079	.551	-.033	.803	.251	.053	-.191	.143
Modified output task 2	55	-.020	.884	-.010	.945	.125	.362	-.237	.082
Modified output both tasks	60	-.089	.498	-.018	.891	.238	.067	-.207	.112

Research Question 5

The fifth research question asked whether aptitude correlated with the effectiveness of recasts in leading to knowledge of the target feature irrespective of the mode of interaction. Table 50 shows that there was no correlation between aptitude and L2 gains when the analysis involved both treatment conditions as one group.

TABLE 50 CORRELATIONS BETWEEN APTITUDE AND L2 DEVELOPMENT FOR BOTH CONDITIONS

Outcome measure (N=60)	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
	rho	p	rho	p	rho	p	rho	p
Oral production test	.066	.617	-.111	.400	.002	.990	-.002	.990
Written production test	.020	.881	-.093	.479	.018	.893	-.055	.674

Research Question 6

The sixth research question asked whether mode of interaction influenced the relationship between L2 aptitude and modified output. Table 51 demonstrates the findings for the FTF condition and Table 52 for the SCMC condition. No correlation was found between L2 aptitude and modified output, neither for the FTF nor for the SCMC group. Similar to other research questions related to learners' modified output, fewer than 30 students made errors in the target feature while performing the second tasks of both conditions. Hence, no recasts were provided and no output modification occurred for these participants.

TABLE 51 CORRELATIONS BETWEEN L2 APTITUDE AND MODIFIED OUTPUT (FTF CONDITION)

FTF	*N*	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
		rho	p	rho	p	rho	p	rho	p
Modified output task 1	30	.022	.908	-.377	.040	-.010	.958	-.002	.991
Modified output task 2	27	.081	.688	-.179	.371	-.017	.935	-.173	.389
Modified output both tasks	30	.009	.963	-.334	.071	.008	.965	-.084	.658

TABLE 52 CORRELATIONS BETWEEN L2 APTITUDE AND MODIFIED OUTPUT (SCMC CONDITION)

SCMC	*N*	*LLAMA B*		*LLAMA D*		*LLAMA E*		*LLAMA F*	
		rho	p	rho	p	rho	p	rho	p
Modified output task 1	30	-.179	.344	.127	.504	.352	.056	-.359	.052
Modified output task 2	28	-.076	.702	.126	.523	.105	.594	-.319	.098
Modified output both tasks	30	-.180	.340	.128	.500	.354	.055	-.356	.054

Research Question 7

The seventh research question asked whether mode of interaction influenced the relationship between L2 aptitude and development in knowledge of the present third person singular. Tables 53 and 54 show the correlations for each group. No relationship was found between aptitude and L2 development, for neither FTF nor SCMC mode.

TABLE 53 CORRELATIONS BETWEEN L2 DEVELOPMENT AND L2 APTITUDE (FTF MODE)

	LLAMA B		*LLAMA D*		*LLAMA E*		*LLAMA F*	
	rho	p	rho	p	rho	p	rho	p
Oral production test	-.009	.964	-.147	.438	-.115	.546	-.051	.790
Written production test	-.099	.602	-.161	.395	-.043	.820	-.082	.667

TABLE 54 CORRELATIONS BETWEEN L2 DEVELOPMENT AND L2 APTITUDE (SCMC MODE)

	LLAMA B		*LLAMA D*		*LLAMA E*		*LLAMA F*	
	rho	p	rho	p	rho	p	rho	p
Oral production test	.255	.174	-.184	.331	.016	.932	.074	.697
Written production test	.290	.119	-.098	.608	-.026	.892	-.073	.700

Summary of Results

In summary, learners in the FTF condition produced significantly more modified output than those in the SCMC condition. Moreover, the FTF group demonstrated significantly greater L2 development than the SCMC group on both oral and written production tests. Only the oral modified output generated by the FTF interaction correlated positively with subsequent L2 development. In contrast, written modified output produced in the SCMC environment was not associated with learners' improvement. Finally, there was no relationship between aptitude and modified output or L2 development in either condition.

4.4.11 Discussion

Research Question 1

The first research question examined the impact of mode of interaction on the amount of modified output produced after recasts targeting the present third person singular. The hypothesis that the SCMC condition would lead to more modified output than the FTF condition was not confirmed. On the contrary, the SCMC group produced a significantly lower amount of modified output than the FTF group. In particular, the mean percentage of modified output was 12.77% in the SCMC condition, and 49.16% in the FTF condition. That is, in SCMC mode only a few participants corrected their initial errors after receiving feedback. This finding is in line with Smith (2010) and Gurzynski-Weiss and Baralt (2014b), where the participants also exhibited low levels of modified output in the SCMC environment. Similar to the present study, Gurzynski-Weiss and Baralt compared the two modes and found that interaction in the FTF mode generated more uptake with modified output as compared to the SCMC environment.

With regard to the current thesis, in order to elucidate possible reasons for the low rate of modified output during online interaction, it is worthwhile to identify different types of responses to recasts that did not involve output modification in the SCMC condition. Five types of responses were elicited: (1) the participants replied “yes”, (2) topic continuation, (3) they repeated the same error, (4) they repeated the same information they typed prior to recasts without correcting their initial errors or (5) they provided additional information related to the semantic content of the pictures of the task without modifying their erroneous utterances. Examples for each type are provided below. In particular, 4.1 corresponds to the first type (i.e. a “yes” answer), 4.2 is an example of topic continuation, 4.3 has repetition of the initial error, in 4.4 the participant simply repeated the same information given prior to feedback (i.e. what the character reads) rather than correct the present third person singular and finally, in examples 4.5 to 4.9, the participants responded to the recasts by offering additional information about the activities of the fictional characters (e.g. their exact job, location, type of dance etc.). Table 55 shows descriptive statistics for the types of responses without modified output the two groups produced.

EXAMPLES OF RESPONSES PROVIDED BY THE SCMC GROUP

(4.1)

R. Lucy?

S. on Saturdays 7 9 is dance not at home

R. She dances?

S. yes

(4.2)

S. mary on Sundays all day is walking

R. She walks?

S. mark on Sundays 6-7 is playing the gutar (guitar) (topic continuation)

(4.3)

S. Is walking.

R. She walks?

S. walking. (same error)

(4.4)

R. Natalie?

S. On Sundays reading book at 3-5 ‘clock at home.

R. She reads?

S. book (same information as prior to the recast)

(4.5)

R. Alex?

S. alex on Saturdays 9-11 is at a home and work

R. He works?

S. acountent (accountant) (additional information)

(4.6)

R. Sophie?

S. Sophie make carts (cards) on Sundays 11-12 a.m. at home.

R. She makes?

S. Carts (cards) with flowers. (additional information)

(4.7)

R. Mary?

S. Mary go for hike on Sundays all day not at home.

R. She goes?

S. A mouten (mountain) (additional information)

(4.8)

R. Jessica?

S. On Saturdays is drawing at 10-12 o'clock at home.

R. She draws?

S. People. (additional information)

(4.9)

R. Lucy?

S. On Saturdays is dancing at 7-9 o'clock not at home.

R. She dances?

S. Tago (tango) (additional information)

TABLE 55 TYPES OF RESPONSES NOT INVOLVING MODIFIED OUTPUT (MO) PER GROUP

TYPE OF RESPONSE	*GROUP (N=30)*	*Mean (%)*	*SD*	*Median*	*IQR*
No MO – Provision of Same Information	FTF	.62	1.91	.00	.00
	SCMC	14.18	20.52	3.57	22.65
No MO – Provision of Novel Information	FTF	1.97	4.47	.00	.86
	SCMC	19.68	29.25	3.38	38.61
Yes Answers	FTF	41.68	37.40	27.71	75.61
	SCMC	35.70	43.07	.00	84.15
Silence and/or Topic Continuation	FTF	6.01	18.19	.00	.83
	SCMC	14.11	33.02	.00	4.42
Same error	FTF	1.44	3.27	.00	.00
	SCMC	1.45	4.39	.00	.00

Note: A few responses that did not involve MO were coded as unclear and thus excluded.

There were no significant differences between the FTF and SCMC groups regarding the first type, that involved "yes" answers (Mann-Whitney U = 361.00, p = .178). The effect size was r = .17. This type of response was also the most frequent for both groups (41.68% for the FTF group, 35.70% for the SCMC group). As explained earlier, these responses may have been encouraged by the interrogative mode of the recasts and do not provide evidence about whether the learners noticed the target feature. There were also no significant differences between the two groups regarding silence/ topic continuation and responses that involved the same error (Mann-Whitney U = 412.50, p = .476 for topic continuation; Mann-Whitney U =

443.00, $p = .882$ for same errors). The effect sizes were $r = .09$ and $r = .01$, respectively.

Interestingly, there were significant differences between the SCMC and FTF groups with respect to other types of responses. In particular, the SCMC group responded to recasts by repeating the same information given prior to feedback (example 4.4) or by offering novel information about the habits of the characters without modifying their non-target-like output (examples 4.5 to 4.9) to a greater extent as compared to the FTF group. These differences were found to be significant (Mann-Whitney $U = 237.00$, $p < .001$ for responses involving same information; Mann-Whitney $U = 280.50$, $p = .005$ for responses providing additional information). The effect sizes were $r = .49$ and $r = .36$, respectively.

Responses involving repetition of the same information given prior to feedback or the provision of additional information might indicate that many of the participants in the SCMC group attended to the semantic content of their utterances rather than to errors related to morphology. As Schmidt (2001) has argued noticing is "nearly isomorphic with attention” (p.1). Thus, not allocating enough attentional resources to the target feature might be related to a lack of noticing of the corrective intention of the recasts and, consequently, of their linguistic target. Nonetheless, it should be emphasized that an interpretation of the results that equates the low rate of modified output with a lack of noticing should be made with caution. First, as explained earlier, modified output is not a reliable predictor of noticing and subsequent development; previous studies have demonstrated contradictory findings about their relationship (Loewen & Philp, 2006; Mackey & Philp, 1998 vs Egi, 2010; Loewen, 2005). Second, several researchers have rightly argued that modified output is a discourse phenomenon not necessarily associated with psycholinguistic processes such as noticing and attention (Long, 2007; Ohta, 2000). Hence, relying solely on modified output to draw conclusions about noticing and L2 outcomes can be highly misleading. In order to gain more information about mode of interaction and noticing, it was examined whether the two groups were different in the numbers of recasts they needed during the treatment, i.e. the number of errors they made in their use of the target construction. According to Table 56, the SCMC group needed more recasts in comparison to the FTF group, and this difference was found to be significant (Mann-Whitney U = 266.50, p = 006). The effect size was .35. This finding indicates that the SCMC group persisted with the same error, requiring the provision of more feedback, whereas the FTF group noticed the recasts and, consequently, they managed to improve their performance during the tasks. As a result, fewer recasts were delivered.

TABLE 56 NUMBER OF RECASTS PER GROUP

GROUP	*NUMBER*	*MEAN*	*SD*	*MEDIAN*	*IQR*
FTF	496	16.53	8.91	16.50	19.00
SCMC	680	22.67	9.98	28.00	17.00

In addition to the above observations, it should also be noted that some learners' practices during SCMC interaction may also demonstrate a lack of noticing. First, some participants' private speech recorded by the screen-capture software involved comments associated with the semantic content of their utterances rather than with the target feature. Second, the text-based online chat recorded by the screen-capture software revealed that most learners did not go back to previous CF episodes to revise messages involving the target linguistic model of the recasts.

Based on the above evidence, if we tentatively accept that there was a lack of noticing and attention to the target feature during SCMC interaction, the question that arises is why the online environment did not facilitate the use of these cognitive mechanisms. One explanation might be related to the target feature, which was a non-salient, communicatively redundant bound morpheme. Although the salience of the target construction may have been increased by the SCMC mode, it remained a redundant feature. Drawing on the SEEV model of selective attention, Wickens (2007) explains that we tend to allocate greater attention to stimuli that we consider more valuable.

> Given the task goal of understanding the semantic meaning of a sentence, there is no doubt that certain words or morphemes can be asserted to be more critical (valuable) to that understanding than others. (Wickens, 2007, 182)

Hence, learners may attend to a greater extent to linguistic areas that they view as more valuable due to their greater contribution to the meaning of their utterances. In line with this argument, previous research has revealed that learners report less noticing for feedback addressing morphology in comparison to other types of errors (e.g. lexical, semantic or phonological errors) (Gurzynski-Weiss & Baralt, 2014b; Kim & Han, 2007; Mackey et al., 2000; Smith, 2012; Trofimovich et al., 2007). Based on these findings, it could be argued that learners might devote greater attention to lexis and the semantic content of their utterances rather than to redundant morphology, such as the present third person singular.

Nevertheless, the above explanation might not paint a complete picture of the findings of the current study as although the same type of morphological recasts targeting the present third person singular were employed in both modes, more modified output was produced in FTF mode than in SCMC mode. Moreover, the students needed more recasts in the SCMC condition as compared to the FTF environment. A possible reason might be related to the participants' expectations during interaction. Learners in the Greek EFL context are used to receiving CF on morphology during FTF interaction, given that error correction in the classroom often occurs in FTF mode. But they are not familiar with the provision of CF while communicating in SCMC mode. Many of them might utilize the SCMC environment in real life rather than in the EFL classroom in order to exchange messages, in which case semantic content is more important than morphological accuracy. Hence, their previous experiences in use of the

SCMC mode might have encouraged them to attend to meaning rather than to form; and consequently, when receiving a recast, many of them simply repeated the same information in order to confirm the semantic content of their utterance or gave additional information because they might have perceived the meaning of their precedent utterance as incomplete.

Another possible explanation for the lower amount of modified output in SCMC mode as compared to FTF mode might be associated with the time required to type a new sentence. That is, some learners might have noticed the feedback in the SCMC mode but did not type the target-like construction in order not to procrastinate during communication. In contrast, uttering the target-like feature after receiving oral feedback during FTF interaction was faster and did not affect the pace of communication. Thus, oral interaction may promote output modification to a greater degree than text-based, online chat.

Research Question 2

The second research question investigated the combined effects of recasts and mode of interaction on knowledge of the present third person singular. The hypothesis that the SCMC environment would be more successful in leading to L2 benefits in comparison to the FTF mode was not confirmed. Unlike this prediction, the study revealed that the FTF group exhibited significantly greater L2 gains than the SCMC group on both outcome measures. Specifically, on the oral production test, the gain score was 27.35% in the FTF condition and only 9.71% in the SCMC condition; and on the written production test, the gain score for the FTF group was 49.24%, whereas for the SCMC group it was only 11.88%. Hence, the FTF environment promoted greater L2 development than the SCMC environment. Interestingly, the students engaging in oral interaction in the FTF mode transferred their knowledge in written mode, as demonstrated by the written production test. In contrast, those performing tasks in SCMC mode did not achieve much improvement in oral mode as revealed by the oral production test. Surprisingly, the SCMC group exhibited only small L2 benefits, even on the written production test, which was administered in the same mode as the treatment.

The difference in the gain scores of the SCMC and FTF groups could be explained in tandem with the amount of modified output they produced. In particular, as explained earlier, the FTF condition generated significantly more modified output and L2 gains than the SCMC condition. Thus, without ignoring the methodological weaknesses of using modified output as a measure of noticing, considering both the low rate of modified output and the limited L2 development achieved in the SCMC environment, it could tentatively be argued that, unlike the FTF group, the SCMC group may not have noticed the target feature during online interaction and, consequently, did not improve their scores. This interpretation is based on Schmidt's (1990, 1995, 2001) argument that noticing linguistic elements in the input is a prerequisite for interlanguage development (see section 2.1.2). The pivotal role of noticing in the SCMC condition has also been demonstrated by Smith

(2012). As mentioned in section 2.2.2, using stimulated recalls and eye-tracking data, Smith showed that noticing linguistic features during SCMC interaction was a predictor of learners' improvement after the provision of recasts. In other words, learners were more likely to achieve L2 gains for linguistic constructions they had noticed.

Regarding the current study, the Limited Capacity Model (Skehan, 2009, 2014b) might offer interesting insights into the reasons why written recasts in SCMC mode did not facilitate L2 development. As explained earlier, many participants in the SCMC group responded to recasts by giving additional information or repeating the same information about the characters' habits. It is possible that due to attentional limitations, the learners could not simultaneously process the target morpheme of the present third person singular and the content depicted in the pictures. Many of their responses to the feedback indicate that they might have devoted more attention to the semantic content of their utterances, rather than to the target construction. It is also possible that, apart from the competition between content and accuracy in the allocation of attentional resources, the learners' attempts to avoid potential orthographic errors might have also increased the cognitive load during written interaction.

The low amount of interlanguage development in the SCMC condition could also be related to the low salience of the present third person singular and to the type of CF. First, with regard to salience, the findings of the present study are in line with Yilmaz (2012), who showed that recasts addressing non-salient linguistic constructions were less effective in leading to development than those targeting salient linguistic features. Second, with respect to the type of feedback, the current results suggest that more explicit interventions may be needed to draw learners' attention to language in the SCMC environment (e.g. explicit feedback, enhanced recasts). Drawing on previous literature, three studies exploring feedback in the SCMC mode have demonstrated that recasts were less beneficial than explicit correction (Yilmaz, 2012) or led to similar learning outcomes as in the control condition where no feedback was provided (Loewen & Erlam, 2006; Sauro, 2009).

Nonetheless, as explained in the discussion section of the first research question, when the two modes are compared (FTF vs SCMC), although the same type of recasts was utilized in both conditions for the same target construction, recasts in the FTF environment were more effective than those in SCMC mode. This may be attributed to the participants' prior learning experiences. In educational contexts, like the one of the study, whereby learners are not familiar with the SCMC mode being utilized as a tool of error correction, recasts addressing morphology may not be successful in drawing learners' attention to linguistic targets. As a result, the linguistic target of written recasts in the SCMC condition remained unattended to. Furthermore, the interrogative mode of the recasts in combination with the learners' prior practice of exchanging messages in a text-based chat where the main focus is usually on meaning rather than on form might have encouraged the SCMC participants to perceive the recasts as clarification requests requiring information about the meaning of their utterances. Hence, recasts in the online environment failed to serve as CF on morphology. Normally, oral recasts in the FTF mode are widely employed by teachers in EFL settings in order to address non-target-like morphological elements. Thus, the FTF mode might have increased the explicitness of the recasts and, consequently, their corrective intention might have been more easily recognized by the EFL learners due to their prior learning experiences. Nicholas, Lightbown, and Spada (2001) have argued that a prerequisite for the effectiveness of recasts leading to L2 benefits is that their intention to correct a form is unambiguous and clear to learners.

Research Question 3

The third research question explored the impact of mode of interaction on the relationship between modified output and development in knowledge of the present third person singular. Considering that previous studies have demonstrated contradictory results with regard to the relationship between modified output and L2 development, no hypothesis was formulated for the third research question. As explained in the section with the results, the study showed that modified output correlated positively with subsequent L2 gains only in FTF mode. This finding suggests that in the FTF condition, when modifying their output, the students possibly employed their internal resources (e.g. working memory, attention) while processing the target feature. An obvious explanation for the lack of correlation between SCMC modified output and L2 development is the low variability of output modification in the SCMC group (IQR= 6.64). Although the majority of the SCMC students did not produce modified output, it is worthwhile to identify the type of modified output the SCMC mode generated. Inspired by Gurzynski-Weiss and Baralt (2014a), output modification was classified into two categories: partial and full modified output. In their study, Gurzynski-Weiss and Baralt coded full modified output as "learners' accurate use of all feedback provided by the interlocutor in the turn immediately following feedback" (p. 1404). Conversely, partial modified output meant that "learners isolated and repeated only the element that had been corrected in feedback" (p. 1404). Gurzynski-Weiss and Baralt found that in the FTF condition the type of modified output (partial vs full) was not associated with the accurate noticing the learners reported during stimulated recalls. On the contrary, the learners reported more noticing of the target item in the SCMC environment when partial modified output was produced. The researchers explain that while producing partial modified output in the SCMC condition, the participants identified and decoded the target feature of the feedback.

In a similar vein, the current study distinguished partial from full modified output. The former was coded as the production of only a personal pronoun and the target-like feature, as in examples 4.10 to 4.12, below. In contrast, full modified output involved output modification followed by the provision of (1) either the same information produced prior to the recast, as in example 4.13, or (2) additional information in order to complete the content of the preceding utterance. In examples 4.14 to 4.16, below, the participants corrected their initial errors related to the present third person singular after receiving recasts; however, they also added novel information with regard to the habits of the characters. In particular, after the target verb, the students added an object in order to give more information about the activity (e.g. what food the character cooks in 4.14, what type of programme the character watches in 4.15 and what the character paints in 4.16).

EXAMPLES OF TYPES OF MODIFIED OUTPUT

(4.10)

R. What about Helen?

S. Helen is at home at 9 to 11 11 o'clock on Saturdays and she watching TV

R. She watches?

S.Ναι (Yes) she watches *(partial modified output, only the target feature)*

(4.11)

R. Mhm. What about Mary?

S. Mary on Satur- on Saturdays in all day not at home climb on mountain

R. She climbs?

S. She climbs ναι (yes) *(partial modified output, only the target feature)*

(4.12)

R. OK. What about Lucy?

S. The Lucy on Saturdays at 5 to 9 is he she dance

R. She dances?

S. She dances *(partial modified output, only the target feature)*

(4.13)

S. Greg study Maths

R. He studies?

S. He studies Maths *(modified output, plus repetition of information provided prior to the recast).*

(4.14)

R. What about George?

S. goerg on Saturdays cook at 9.00 at home

R. He cooks?

S. he cooks a salat (salad) *(modified output, plus additional information)*

(4.15)

R. Tell me about Helen.

S. Helen on Saturdays is at home from 9-11 she watch tv in her sofa

R. She watches?

S. I think she watches documentaries *(modified output, plus additional information)*

(4.16)

R. Nick?

S. nick on sandays painting at 4-6 at home sundays

R. He paints?

S. He paints patterns *(modified output, plus additional information)*

After identifying and measuring full and partial modified output, the two modes were compared in terms of the amount of partial modified output they produced, as this type of modification may be related to decoding and more deeply processing the linguistic target of the recasts. Table 57 shows descriptive statistics for partial modified output per group.

TABLE 57 PARTIAL MODIFIED OUTPUT PER GROUP

TYPE OF MODIFIED OUTPUT	*GROUP (N=30)*	*Mean (%)*	*(SD)*	*Median*	*IQR*
Partial Modified Output	FTF	18.61	(18.29)	15.47	27.19
	SCMC	.62	(2.38)	.00	.00

Interestingly, partial modified output was mainly observed in the FTF group, and it was very limited in the SCMC group. The difference between the two groups was significant (Mann-Whitney U = 119.00, $p = < .001$), and the effect size was large, $r = .70$. It could be argued that production of only the target verb after receiving feedback in the FTF group might have led to the allocation of more cognitive resources to the present third person singular, and hence modified output correlated positively with subsequent development. In line with Gurzynski-Weiss and Baralt, the present study shows that when modified output is part of a longer utterance involving more information, this might not encourage learners to identify and decode the target construction in order to engage in a cognitive comparison between their error and the target-like item. The semantic content of a long utterance may outweigh the target construction and obscure the linguistic area addressed by the feedback. Hence, the tendency of some students in the SCMC group to give more details about the characters' activities when modifying their output may have encouraged them to allocate more attention to these details, rather than to the present third person singular. In contrast, when learners modify only the non-target-like verb, they might process the target feature more deeply. That is to say, partial modified output in FTF mode might have relieved the cognitive burden of oral interaction, thus enabling the use of greater memory and attentional resources to be directed towards processing of the target construction.

It is also worth noting that although modified output was not associated with L2 gains in the SCMC condition, production of the present third person singular in new verbs after receiving recasts whilst engaging with treatment tasks correlated positively with their subsequent development. Put differently, those students who produced the target construction accurately in novel utterances during the treatment, after the provision of one or more recasts, exhibited L2 benefits on both the oral and written production tests (rho = .575, p =.001 for the oral production test; rho = .624, $p < .001$ for the written production test). Hence, in SCMC mode, although immediate uptake with output modification was not an indicator of learning, production of the target construction in new verbs revealed the learners' ability to generalize the target pattern and exhibit greater improvements on the tests. In FTF mode, correct production of the target feature in new verbs during the treatment was also associated with the participants' development (rho =.580, p =.001 for the oral production test; rho = .495, p =.005 for the written production test). Thus, the ability to use the target feature in novel utterances is a reliable demonstrator of L2 outcomes in different experimental conditions.

Research Questions 4-7

The remaining research questions aimed to provide insights into the role of aptitude. In particular, the fourth and fifth research questions investigated whether there was a relationship between (1) aptitude and modified output and (2) aptitude and L2 development when the FTF and SCMC conditions were analysed as one group (N=60). The study demonstrated that L2 aptitude was not associated with either modified output or L2 gains. Hence, the next step was to examine the role of mode of interaction. To this end, the sixth and seventh research questions were formulated. Their aim was to examine whether mode of interaction influenced potential correlations of L2 aptitude with modified output and L2 gains. The hypothesis was that low aptitude students would benefit from written feedback provided in SCMC mode as this environment may make the present third person singular more salient and students would have more time (1) to process the input when delivered in written rather than oral mode and (2) to engage in online planning (Ellis, 2005b; Yuan & Ellis, 2003). Counter to these predictions, the study showed that in neither condition (FTF and SCMC) was aptitude associated with modified output or L2 development.

The lack of a relationship between aptitude and the rate of modified output in the FTF and SCMC conditions cannot provide robust evidence about whether learners equipped with greater aptitude notice the feedback as the production of modified output is not obligatory. That is, learners with high scores on the LLAMA aptitude test might have noticed the target construction during interaction without modifying their erroneous utterances. As for learners' development, surprisingly, none of the components of L2 aptitude examined by the present study correlated with L2 gains in either the FTF or SCMC condition. However, this finding may be interpreted differently in the two modes. In the FTF condition, the learners improved their scores regardless of L2 aptitude, whereas in the SCMC condition the learners achieved very limited gains irrespective of their aptitude scores. In other words, in the FTF group, the effects of L2 aptitude were mitigated and oral recasts benefited learners with both low and high aptitude. By contrast, in SCMC mode, high aptitude did not facilitate L2 development. The interquartile range in SCMC mode was also relatively low (12.71 on the oral production test, 3.43 on the written production test). Low variability indicates that there were no marked differences among the students in their L2 development. Low variability can also explain the lack of correlations between L2 gains and aptitude.

A question that arises is why the learners in the SCMC group did not exploit their higher aptitude (e.g. analytic ability, ability to associate sounds with symbols) in order to benefit from the written recasts. A possible explanation is that even learners with great cognitive ability allocated more attention to the content of the information they produced rather than to the formal aspects of the second language. Moreover, according to previous literature, L2 aptitude seems to be less facilitative of non-salient linguistic constructions; and regarding analytic ability, it has been found to contribute to L2 learning when more explicit feedback is provided (Li, 2013; Sheen, 2007; Yilmaz, 2013; Yilmaz & Granena, 2015). Thus, using recasts rather than more explicit feedback to address a non-salient linguistic feature may have negatively influenced the facilitative role of language analytic ability.

As for the FTF condition, a possible reason why the participants exhibited improvement regardless of aptitude might be associated with the combined effects of (1) the low cognitive demands of the information transmission task and (2) the learners' familiarity with the mode. With regard to the former, as explained in study 1 (chapter 3), the participants judged that the information transmission tasks required a relatively low amount of mental effort. Tasks with low cognitive demands might be beneficial for L2 learners, especially in the first stages of acquiring a linguistic construction, as they might allow them to devote greater attention to the target feature. Nonetheless, the same tasks did not facilitate development in SCMC mode. First, a task that is judged simple in FTF mode does not necessarily imply that it imposes low cognitive demands in SCMC mode. Baralt (2013) showed that mode of interaction influenced whether the participants perceived a task as simple or complex in a post-task perception questionnaire. In Baralt, FTF groups considered their tasks more demanding than CMC groups, regardless of whether they carried out simple or complex versions of the task. In the current study, the young learners' efforts to provide additional information without making orthographical errors in SCMC mode may have increased the cognitive demands of tasks judged as simple in FTF mode. Nonetheless, a post-task questionnaire was not administered to the SCMC group, and hence conclusions cannot be drawn about whether the information transmission tasks were perceived as more complex in SCMC mode.

With respect to mode familiarity, it refers to the participants' previous learning experiences. As noted earlier in this chapter, the FTF environment might serve as a more favourable condition for error correction when it involves EFL learners who may expect to receive CF as a result of previous learning routines. Hence, even low aptitude students may have been predisposed towards CF when engaging in oral interaction. By contrast, text-based online interaction may not have been perceived by students as a learning environment that would involve correction of morphology. Consequently, they did not exploit their high aptitude in language analytic ability and their ability to figure out sound-symbol correspondences so as to improve their scores on the oral and written production tests.

An important observation that should be emphasized is that although the FTF mode induced greater development than the SCMC mode, and its benefits were not influenced by aptitude constructs measured by the LLAMA test, great variability was found in the gain scores of the FTF group. The interquartile range was 53.04 on the oral production test, and 100.00 on the written production test. This indicates that although some students made considerable gains, others did not. For example, on the written production test, some learners did not produce the target feature even once on the pretest, scoring 0%, while they produced it in all of the obligatory contexts on the posttest, scoring 100%. By contrast, other students in the same group demonstrated no development on the written production test, as they scored 0% on both pretest and posttest. This suggests that other variables may affect learners' development, such as learners' motivation, anxiety levels and/or aptitude constructs not measured by the current study.

CHAPTER 5: CONCLUSION

The results and discussion of the two studies presented in chapters 3 and 4 lead to several conclusions that have both theoretical and pedagogical implications. This chapter provides a summary of the two studies, discusses the conclusions, highlights the limitations of the current thesis and, finally, offers suggestions for future research.

5.1 Study 1

5.1.1 Summary

Study 1 sought to investigate the influence of task complexity on the amount of modified output produced after recasts, the combined effects of task complexity and recasts on facilitating knowledge of the present third person singular, and whether task complexity affected the relationship between modified output and development while recasts were held constant. It also uncovered relationships between aptitude and modified output and L2 development, and whether these relationships were influenced by task complexity.

The study demonstrated that although the simple and complex conditions did not differ in the amount of modified output generated by recasts, learners engaging in simple tasks benefitted to a greater extent from recasts than those performing complex tasks. Moreover, a positive relationship was found between modified output and L2 development measured by an oral production test in both conditions, and measured by a written production test only in the simple condition. Finally, aptitude was associated with neither modified output nor gain scores when the two conditions were analysed as one group. When the two groups were examined separately, aptitude did not correlate with modified output; however, it was linked with L2 development in the complex condition. Interestingly, the aptitude constructs that were related to learners' gains were their ability to recognize oral patterns measured by LLAMA D and their ability to associate sounds with their symbols measured by LLAMA E. The above findings have certain implications both at a theoretical level in the field of SLA, and at a pedagogical level in the ELT classroom. These implications are discussed below.

5.1.2 Theoretical Implications

Study 1 demonstrated that learners receiving recasts during simple tasks achieved higher gain scores than those under the complex condition. This result is in line with Skehan's (2009, 2014b) suggestion that when a task eases the processing demands in the conceptualizer stage, more attentional resources are available for processing linguistic constructions. In Study 1, recasts were more effective in drawing learners' attention to the linguistic target when they were provided during simple tasks that did not require great mental effort; and they did not consume learners' attentional resources in the conceptualizer stage while processing the content of the task. On the contrary, learners performing tasks under the complex condition may have devoted more attention to the content of their utterances when making decisions about the owners of items (see section 3.4.4); and as a result, they did not process the target feature at a level that would enable them to improve their performance during oral and written production.

The results of the current thesis also confirm Robinson's (2011) suggestion that there is greater variation in L2 outcomes when learners perform complex tasks as a result of individual differences in several constructs such as learners' aptitude. When the simple and complex conditions were examined as one group, aptitude was not associated with L2 gains. But when the two groups were investigated separately, aptitude was related to L2 development only in the complex condition, while it was not linked to the gain scores of students in the simple condition. This result indicates that aptitude may play a crucial role under certain experimental conditions, while its impact is eliminated under others. Hence, research on aptitude-treatment interactions can provide interesting insights into the extent to which aptitude moderates the benefits of certain interventions. It should also be underlined that aptitude is not a holistic construct as it involves different components. As explained earlier (section 2.4.6), Robinson (2001a, 2002, 2007a, 2012) has proposed that different aptitude constructs may be related to different learning conditions. For example, in the complex condition of the current study, when the students received oral recasts, the components of aptitude that were relevant to learners' improvement were their ability to recognize oral patterns and to associate sounds with symbols (phonetic coding).

5.1.3 Pedagogical Implications

Two main pedagogical implications can be inferred from Study 1. First, at least under complex task conditions, young learners with low aptitude in recognizing sounds and in associating sounds with symbols may not benefit from oral recasts targeting allomorphs of the present third person singular. In other words, these allomorphs, which are sounds added to the base form of the verb, may not be amenable to recasts delivered in oral mode for students with low aptitude scores in oral pattern recognition and phonetic coding.

A second implication is that engaging with cognitively simple tasks may be more beneficial for young learners. The study showed that when learners received recasts while performing tasks that did not require great mental effort, they were more likely to engage in cognitive comparisons of their erroneous utterances and the target-like feature, regardless of their aptitude. Furthermore, oral recasts supplied during simple tasks enabled learners to develop declarative knowledge, and hence improve their scores on the written production test, as they had more time to think about the underlying pattern related to the target construction.

5.2 Study 2

5.2.1 Summary

Apart from task complexity, another factor that may influence the efficacy of recasts is whether interaction occurs in the FTF or SCMC mode. In order to shed light on this variable, study 2 examined the impact of mode of interaction on the rate of modified output produced after recasts, the combined effects of mode of interaction and recasts on promoting knowledge of the present third person singular, and whether mode of interaction influenced the relationship between modified output and L2 development while recasts remained constant. It also investigated the relationships between aptitude and modified output and L2 gains, and whether these relationships were affected by the mode of interaction.

The study found that recasts delivered in FTF mode generated significantly more modified output than those supplied in SCMC mode. Moreover, the FTF group achieved significantly greater L2 gains than the SCMC group on both oral and written production tests. Modified output was associated with learners' improvement only in the FTF condition. Finally, aptitude was not associated with L2 benefits in either the FTF or SCMC environment. The theoretical and pedagogical implications of these results are presented below.

5.2.2 Theoretical Implications

The findings of the present study appear to provide support for the Limited Capacity Model (Skehan, 2009, 2014b). This model seems to explain why SCMC recasts did not lead to marked L2 benefits. As explained in the discussion part of the fourth chapter, many of the participants gave additional information related to the content of the pictures after receiving morphological recasts. This may indicate that they devoted more attention to the semantic content of their utterances, at the expense of grammar. Although it was expected that the written mode of the recasts would relieve the cognitive burden for the learners due to the lower pace of interaction and the possibility of online planning (Ellis, 2005b; Yuan & Ellis, 2003), the SCMC mode had the opposite effect. As the learners were unfamiliar with the mode being utilized as a learning environment focusing on morphology, they devoted more attention to the meanings of their utterances. As their responses to the feedback reveal (see section 4.4.11), after receiving morphological recasts, the participants attempted to make the meanings of their utterances more complete by providing additional information or repeating the same information they had given prior to feedback. In other words, they prioritized meaning rather than form-meaning mappings. Hence, drawing on the Limited Capacity Model (see the theoretical review in section 2.3.1), it seems that the SCMC mode increased the cognitive load in the conceptualizer stage (i.e. processing the content of pictures and the meaning of messages). Such a cognitive burden probably resulted in less attention being devoted to linguistic encoding and to the target construction. Furthermore, the attempts by some participants not to make orthographical errors may have augmented the cognitive burden of the tasks in SCMC mode to a greater extent in comparison to the FTF mode.

5.2.3 Pedagogical Implications

As explained earlier, drawing on previous literature, the SCMC mode was expected to be beneficial to L2 learners for several reasons: First, it was assumed that the SCMC environment would increase the salience of the bound morphemes of the present third person singular and, consequently, facilitate modified output and subsequent L2 development. Moreover, it was predicted that the SCMC mode would benefit both low-aptitude and high-aptitude learners. For example, the potential to plan their output and to go back and revise previous recasts was expected to decrease the cognitive load of interaction, and hence assist those learners with low aptitude scores.

Despite these predictions, the SCMC mode was significantly less effective than the FTF mode in triggering modified output after recasts, and in leading to L2 improvement from pretests to posttests. Interestingly, even learners with high aptitude scores did not exhibit large gains in the SCMC group. It was also observed that many students in the SCMC condition tended to provide additional information regarding the content of the pictures after receiving recasts. This indicates that the participants may have allocated greater attention to meaning at the expense of form.

Therefore, when recasts are provided in SCMC mode, it seems that learners' attention is not drawn to the target construction simply because it is written on a screen and can be revised. That is, the mode itself does not seem to increase the salience of a linguistic feature, at least for young learners in the first developmental stages. Nicholas, Lightbown, and Spada (2001) have rightly argued that L2 learners benefit from recasts when they perceive them as corrective moves targeting linguistic errors. If learners interpret recasts as confirmation checks or clarification requests requiring more details related to meaning, the linguistic target will probably remain unattended to regardless of whether interaction takes place in FTF or SCMC mode. As the EFL learners in the current study may have been more familiar with error correction occurring in FTF mode rather than in SCMC mode, the corrective intention of oral recasts was probably less ambiguous to the students in comparison to written recasts in the SCMC environment. Hence, when teachers use the SCMC mode to provide recasts addressing morphology, they should probably inform the students in advance so as to draw their attention to target linguistic areas.

5.3 Limitations and Directions for Future Research

There are some limitations to the present studies that should be acknowledged, and these should be considered in future research. First, both Studies 1 and 2 compared two experimental conditions in terms of the amount of modified output produced after recasts, and they also examined the relationship of modified output with (1) subsequent L2 gains and (2) participants' aptitude. Nonetheless, as explained earlier, modified output is not a robust demonstrator of noticing. Thus, future research could employ alternative techniques, such as stimulated recalls or immediate reports, to measure the noticing of feedback during simple and complex tasks, or eye-tracking during SCMC interaction in tandem with stimulated recalls to obtain a more direct view of attentional allocation and noticing processes.

A second limitation is that only one type of feedback was utilized: partial, interrogative recasts. A replication of the present thesis could include more explicit CF techniques (e.g. metalinguistic feedback or explicit correction) or more implicit interventions, such as full recasts provided in interrogative or declarative mode. As explained in chapter 2, recasts with different features may differentially affect noticing and learning. Regarding oral feedback in FTF mode, the present thesis demonstrated that the aptitude construct involving sound recognition (i.e. LLAMA D) correlated positively with learners' L2 gains after receiving recasts. In light of this finding, it would be of great pedagogical value if future research explored whether learners with low LLAMA D scores benefit more from more explicit feedback or a combination of oral and written feedback (e.g. writing the construction during oral communication or using phonetic transcription for morphology). With respect to SCMC feedback, research could delve into whether more explicit interventions are successful in promoting L2 gains for young learners. A replication of the current study could also involve questionnaires or interviews in order to gain information about learners' perceptions with regard to the SCMC mode (e.g. whether they view it as a learning environment offering opportunities for improving their grammar or only perceive it as a mode that serves as a tool of communication).

A third shortcoming that should be mentioned is that the current empirical work focused on only a single type of grammatical construction (i.e. the present third person singular) and it cannot be generalized to lexis, phonology or other morphosyntactic features, especially if they differ in perceptual saliency and redundancy. For example, CF delivered during more complex tasks may benefit learners when a linguistic construction is needed for completion of a task (e.g. feedback on discourse markers showing complex relationships between events). Hence, future research should elucidate the role of task complexity and feedback for different linguistic areas. With respect to the second study of the current thesis, it demonstrated that when recasts target redundant morphology in SCMC mode, they are not successful when learners prioritize meaning over form as reflected by their responses to feedback. As shown earlier, these responses involved repetition of the same information or the provision of additional information related to content. Future research could examine whether recasts targeting lexical errors or morphosyntactic features that contribute more to meaning than the present third person singular are more effective in drawing learners' attention to language.

A fourth limitation concerns the construct of task complexity. The present thesis only investigated task complexity in FTF mode and only in accordance with +/-reasoning demands. Future studies should broaden the scope of the present research project, and delve into the combined effects of CF and task complexity addressed by other factors (e.g. +/- few elements) in both FTF and online environments. It would also be interesting if a third condition was examined by future research involving more complex conditions than the decision-making tasks employed in the present thesis. For example, although the post-task questionnaire demonstrated that the learners judged the decision-making tasks to be significantly more cognitively demanding than the information transmission tasks, no difference was found between the two experimental conditions in terms of task difficulty. A possible explanation is that although the students had to think in order to relate the items of the task to potential owners, the decisions they had to make were relatively easy (e.g. to relate a lamp to someone who works at night). Future research projects could have a third experimental condition with decision-making tasks requiring more difficult and less obvious decisions (i.e. simple tasks, +complex tasks, ++complex tasks, as in Kim, 2012).

Regarding aptitude, the present empirical work utilized only a single measure, i.e. the LLAMA test. Future studies should delve more deeply into the relationship between individual differences in cognitive abilities with CF delivered during simple vs complex tasks or in FTF vs SCMC mode. For example, the role of WMC and other aptitude constructs, such as attentional control, should also be explored. For example, the current thesis found that LLAMA D played an important role in the efficacy of oral recasts. The relationship between aptitude in sound recognition as measured by LLAMA D and the effectiveness of oral recasts in facilitating L2 development may imply that phonological WMC is also associated with the extent to which L2 learners process oral feedback successfully. Moreover, although the present empirical work found no relationship between aptitude and L2 development after receiving recasts in simple task conditions, the interquartile range for the information transmission group was large (57.19 and 100.00 for gains on oral and written production tests, respectively). This suggests that other constructs may play a pivotal role in second language development, such as motivation, WMC, the extent to which learners prioritize accuracy in their speech etc. Future studies should also illuminate the relationship between cognitive factors and CF during more spontaneous interaction under natural conditions or in a classroom-based context.

In addition, the present studies employed only tests that required L2 production. Future studies should involve comprehension-based outcome measures, such as grammaticality judgement tests, in order to gauge additional aspects of learners' development. A replication of the current project could also utilize less controlled oral and written production tests so as to assess learners' knowledge during more spontaneous production. Moreover, a delayed post-test should also be administered to provide more evidence about whether learners with high aptitude in phonetic coding and sound recognition maintain their L2 benefits in both oral and written modes.

Finally, the participants in the current thesis were young EFL learners in Greece. Hence, the results cannot be generalized to other age groups or to learners in different educational contexts, at different proficiency levels or with other L1 backgrounds. Future studies could investigate the impact of CF, task complexity, mode of interaction and cognitive variables on the L2 outcomes of adult learners in different instructional contexts and/or in more advanced stages of development.

5.4 Concluding Remarks

Although a plethora of research has explored the effects of CF, and in particular recasts on L2 learning, CF is a complicated construct whose efficacy is linked to several factors. This thesis has revealed that three factors play a pivotal role and merit further research. First, task conditions imposing different cognitive requirements constitute an important variable that influences the extent to which young learners benefit from oral recasts. Second, learners' L2 aptitude, and in particular cognitive individual differences in sound recognition and in phonetic coding, are associated with the processing of oral recasts when performing complex tasks (i.e. tasks requiring greater mental effort). Finally, young learners in some EFL contexts may benefit more from oral, morphological recasts delivered in FTF mode in comparison to written, morphological recasts in SCMC mode. Although the current thesis indicates that these three factors (i.e. L2 aptitude, task complexity and mode of interaction) impact on the effectiveness of recasts, more research addressing the limitations described above is needed to delve more deeply into these variables and their relationships.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. https://doi.org/10.1017/S027226310808073X

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*(1), 35–59. https://doi.org/10.1177/1362168810383329

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit l2 knowledge. *Language Teaching Research*, *16*(1), 9–37. https://doi.org/10.1177/1362168811423339

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, *28*(4), 543–574. https://doi.org/10.1017/S0272263106060268

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, *19*(4), 537–552. https://doi.org/doi:10.1017/S0142716400010353

Baddeley, A. (2000). The epissodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*(4), 689 – 725. https://doi.org/https://doi.org/10.1017/S0272263113000429

Bardovi-Harlig, K. (1987). Markedness and Salience in Second-Language Acquisition. *Language Learning*, *37*(3), 385–407. https://doi.org/10.1111/j.1467-1770.1987.tb00577.x

Bigelow, M., Delmas, R., Hansen, K., & Tarone, E. (2006). Literacy and the processing of oral recasts in SLA. *TESOL Quarterly*, *40*(4), 665–689. https://doi.org/10.2307/40264303

Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *Modern Language Journal*, *93*(2), 227–240. https://doi.org/10.1111/j.1540-4781.2009.00858.x

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*(2), 247–271. https://doi.org/10.1017/S0272263110000756

Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative speaker interactions. *Language Learning*, *52*(1), 1–42. https://doi.org/10.1111/1467-9922.00176

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 134–146). London: Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning teaching and testing* (pp. 23–48). Harlow, UK: Longman.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does First Language Maintenance Hamper Nativelikeness in a Second Language? *Studies in Second Language Acquisition*, *34*(2), 215–241. https://doi.org/10.1017/S0272263112000034

Carroll, J. B. (1965). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training, research, and education* (pp. 87–136). New York: Wiley.

Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Rowley, MA: Newbury House.

Carroll, J. B., & Sapon, S. (2002). *Manual for the MLAT*. Bethesda, MD: Second Language Testing.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test.* San Antonio, TX: Psychological Corporation.

Chaudron, C. (1977). A DESCRIPTIVE MODEL OF DISCOURSE IN THE CORRECTIVE TREATMENT OF LEARNERS' ERRORS. *Language Learning*, *27*(1), 29–46. https://doi.org/10.1111/j.1467-1770.1977.tb00290.x

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology*, *32*(5), 867–873. https://doi.org/10.1037/0012-1649.32.5.867

Chun, D. M. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, *22*(1), 17–31. https://doi.org/10.1016/0346-251X(94)90037-X

Cobb, T. (2016). VocabProfiler [Computer software]. https://doi.org/Retrieved from http://www.lextutor.ca/vp/eng

Corder, S. P. (1967). The significance of learner's errors. *IRAL - International Review of Applied Linguistics in Language Teaching*, *5*(1–4), 161–170. https://doi.org/10.1515/iral.1967.5.1-4.161

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, *11*(4), 367–383. https://doi.org/10.1017/S0272263100008391

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology*, *17*(4), 367–378. https://doi.org/10.1037/0012-1649.17.4.367

de Graaff, R. (1997). The Experanto Experiment. *Studies in Second Language Acquisition*, *19*(2), 249–276. https://doi.org/doi:10.1017/s0272263197002064

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–113). Mahwah, NJ: Erlbaum.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413–438. https://doi.org/doi:10.1017/S0142716410000056

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533. https://doi.org/0272-2631/00

DeKeyser, R. M., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-Linguistic Evidence for the Nature of Age Effects in Second Language Acquisition. *Applied Psycholinguistics*, *31*(3), 413–438. https://doi.org/10.1017/S0142716410000056

DeKeyser, R. M., & Criado, R. (2013). Automatization, skill acquisition, and practice in second language acquisition. In *The Encyclopedia of Applied Linguistics* (pp. 1–8). https://doi.org/10.1002/9781405198431.wbeal0067

Doughty, C. J. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). Cambridge, UK: Cambridge University Press.

Doughty, C. J., & Varela, E. (1998). Communicative focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). New York: Cambridge University Press.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). New York: Cambridge University Press.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. Oxford: Oxford University Press.

Egi, T. (2007a). Interpreting recasts as linguistic evidence: The roles of linguistic target, length, and degree of change. *Studies in Second Language Acquisition*, *29*(4), 511–537. https://doi.org/10.1017/S0272263107070416

Egi, T. (2007b). Recasts, learners’ interpretations, and L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 249–267). Oxford: Oxford University Press.

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *Modern Language Journal*, *94*(1), 1–21. https://doi.org/10.1111/j.1540-4781.2009.00980.x

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*(1), 91–126. https://doi.org/https://doi.org/10.1017/S0272263100014698

Ellis, N. C. (2004). The processes of second language acquisition. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 49–76). Mahwah, NJ: Erlbaum.

Ellis, N. C. (2005). At the interface: {Dynamic} interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305–352. https://doi.org/10.1017/S027226310505014X

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*(2), 145–171. https://doi.org/https://doi.org/10.1017/S0272263197002027

Ellis, N. C., & Sinclair, S. G. (1996). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 234–250. https://doi.org/10.1080/713755604

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2005a). Measuring implicit and explicit knowledge of a second language: A Psychometric Study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2005b). *Planning and task performance in a second language*. Amsterdam: John Benjamins.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, *27*(3), 431–463. https://doi.org/10.1093/applin/aml022

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 339–360). Oxford: Oxford University Press.

Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, *1*(1), 2–18. https://doi.org/10.5811/westjem.2011.5.6700

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner Uptake in Communicative ESL Lessons. *Language Learning*, *51*(2), 281–318. https://doi.org/10.1111/1467-9922.00156

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263106060141

Ellis, R., & Sheen, Y. (2006). Reexamining the Role of Recasts in Second Language Acquisition. *Studies in Second Language Acquisition*, *28*(4), 575–600. https://doi.org/10.1017/S027226310606027X

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171. https://doi.org/10.1191/1362168805lr161oa

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*(3), 464–491. https://doi.org/10.1093/applin/aml001

Erlam, R., & Akakura, M. (2016). New developments in the use of elicited imitation. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 93–111). Routledge.

Foster, P., & Skehan, P. (1996). The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, *18*(3), 299. https://doi.org/10.1017/S0272263100015047

Gallimore, R., & Tharp, R. G. (1981). THE INTERPRETATION OF ELICITED SENTENCE IMITATION IN A STANDARDIZED CONTEXT. *Language Learning*, *31*(2), 369–392. https://doi.org/10.1111/j.1467-1770.1981.tb01390.x

Gass, S. M. (1990). Second and foreign language learning: Same, different or none of the above? In B. VanPatten & J. Lee (Eds.), *Second language acquisition - Foreign language learning* (pp. 34–44). Clevedon: Multilingual Matters.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.

Gass, S. M. (2003). Input and interaction. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 224–255). Oxford: Blackwell.

Gass, S., & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, *49*(4), 549–581. https://doi.org/10.1111/0023-8333.00102

Gathercole, S., & Baddeley, A. (1993). *Working Memory and Language*. Hove, UK: Lawrence Erlbaum Associates.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966–979. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9383619

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *IRAL - International Review of Applied Linguistics in Language Teaching*, *47*(3–4), 367–395. https://doi.org/10.1515/iral.2009.016

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English. A meta-analysis of multiple determinants. *Language Learning*, *51*(1), 1–50. https://doi.org/10.1111/1467-9922.00147

Goo, J. (2012). Corrective Feedback and Working Memory Capacity in Interaction - Driven L2 Corrective Feedback and Working Memory Capacity in Interaction-Driven L2 Learning. *Studies in Second Language Acquisition*, *34*(3), 445–474. https://doi.org/10.1017/S0272263112000149

Goo, J. (2016). Corrective feedback and working memory capacity: A replication. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition.* (pp. 279–302). Amsterdam: John Benjamins.

Goo, J., & Mackey, A. (2013). The Case Against the Case Against Recasts. *Studies in Second Language Acquisition*, *35*(1), 127–165. https://doi.org/10.1017/S0272263112000708

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–130). Amsterdam: John Benjamins.

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics*, *35*(4), 483–503. https://doi.org/10.1093/applin/amu013

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A Theory-Based Approach to the Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test. *The Modern Language Journal*, *84*(3), 390–405. https://doi.org/10.1111/0026-7902.00076

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *56*(7), 1213–1236. https://doi.org/10.1080/02724980343000071

Gurzynski-Weiss, L., & Baralt, M. (2014a). Does type of modified output correspond to learner noticing of feedback ? A closer look in face-to-face and computer-mediated task-based interaction. *Applied Psycholinguistics*, *36*, 1393–1420. https://doi.org/:10.1017/S0142716414000320

Gurzynski-Weiss, L., & Baralt, M. (2014b). Exploring Learner Perception and Use of Task-Based Interactional Feedback in FTF and CMC Modes. *Studies in Second Language Acquisition*, *36*(1), 1–37. https://doi.org/10.1017/S0272263113000363

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL QUARTERLY*, *36*(4), 543–572. https://doi.org/10.2307/3588240

Han, Z. (2013). Forty years later: Updating the fossilization hypothesis. *Language Teaching*, *46*(2), 133–171. https://doi.org/10.1017/S0261444812000511

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*(3), 379–400. https://doi.org/10.1017/S0272263197003045

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(1), 25. https://doi.org/10.1017/S0272263100010457

Havranek, G. (2002). When is corrective feedback most likely to succeed? *International Journal of Educational Research*, *37*(3), 255–270. https://doi.org/10.1016/S0883-0355(03)00004-1

Hawkes, M. L. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal*, *66*(3), 327–336. https://doi.org/https://doi.org/10.1093/elt/ccr059

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range and frequency programs [Computer software]. https://doi.org/Retrieved from http://www.victoria.ac.nz/lals/resources/range.aspx

Hwu, F., & Sun, S. (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System*, *40*(4), 505–521. https://doi.org/10.1016/j.system.2012.10.009

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form -te i-(ru) by learners of Japanese as a foreign language. *Language Learning*, *54*(2), 311–394. https://doi.org/10.1111/j.1467-9922.2004.00257.x

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition*, *25*(1), 1–36. https://doi.org/https://doi.org/10.1017/S0272263103000019

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning*, *51*(September), 401–436. https://doi.org/10.1111/0023-8333.00160

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, *102*, 199–225. https://doi.org/10.1111/j.0079-1636.2004.00135.x

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kartchava, E., & Ammar, A. (2014). The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research*, *18*(4), 428–452. https://doi.org/10.1177/1362168813519373

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 143–164). Philadelphia: John Benjamins.

Keating, G. D., & Jegerski, J. (2014). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*(1), 1 – 32. https://doi.org/https://doi.org/10.1017/S0272263114000187

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*(4), 703–746. https://doi.org/10.1111/j.1467-9922.2008.00477.x

Kern, R. G. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *The Modern Language Journal*, *79*(4), 457–476. https://doi.org/10.1111/j.1540-4781.1995.tb05445.x

Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL*

Journal, *4*(1), 57–72.

Kim, J. H., & Han, Z.-H. (2007). Recasts in communicative EFL classes: Do teacher intent and learner interpretation overlap? In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 269 – 297). Oxford: Oxford University Press.

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition*, *34*(4), 627– 658. https://doi.org/https://doi.org/10.1017/S0272263112000368

Kim, Y. J., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, *41*(3), 829–840. https://doi.org/10.1016/j.system.2013.08.005

Kim, Y., Payant, C., & Pearson, P. (2015). THE INTERSECTION OF TASK-BASED INTERACTION, TASK COMPLEXITY, AND WORKING MEMORY. *Studies in Second Language Acquisition*, *37*(3), 549–581. https://doi.org/10.1017/S0272263114000618

Kormos, J. (2011). Speech production and the cognition hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39–60). Amsterdam: John Benjamins.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

Kuiken, F., & Vedder, I. (2007). Cognitive task complexity and linguistic performance in French L2 writing. In Mayo Maria del Pilar García (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 117–135). Clevedon: Multilingual Matters.

Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity. Researching the cognition hypothesis of language learning and performance.* (pp. 91–104). Amsterdam: John Benjamins.

Lai, C., Fei, F., & Roots, R. (2008). The Contingency of Recasts and Noticing. *CALICO Journal*, *26*(1), 70–90. Retrieved from http://proxy.mul.missouri.edu/login?url=http://search.proquest.com/docview/61992008?accountid=14576%5Cnhttp://ew3dm6nd8c.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQ:ericshell&rft_val_fmt=info:ofi/fmt:ke

Lai, C., & Zhao, Y. (2006). Noticing and text-based chat. *Language Learning & Technology*, *10*(3), 102–120. https://doi.org/http://dx.doi.org/10125/44077.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, *25*(1), 37–63. https://doi.org/https://doi.org/10.1017/S0272263103000020

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, *57*(2), 229–270. https://doi.org/10.1111/j.1467-9922.2007.00408.x

Lenzing, A. (2015). Exploring Regularities and Dynamic Systems in L2 Development. *Language Learning*, *65*(1), 89–122. https://doi.org/10.1111/lang.12092

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, S. (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language Learning*, *60*(2), 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *Modern Language Journal*, *97*(3), 634–654. https://doi.org/10.1111/j.1540-4781.2013.12030.x

Li, S. (2015). Working memory, language analytical ability and L2 recasts. In Z. Wen, M. Borges, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 139–159). Bristol: Multilingual Matters.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/https://doi.org/10.1017/S027226311500042X

Lin, W., Huang, H.-T., & Liou, H.-C. (2013). The effects of text-based SCMC on SLA: A meta analysis. *Language Learning & Technology*, *17*(2), 123–142. Retrieved from http://llt.msu.edu/issues/june2013/linetal.pdf

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., … Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*(3), 530–566. https://doi.org/10.1111/lang.12011

Llinares, A., & Lyster, R. (2014). The influence of context on patterns of corrective feedback and learner uptake: A comparison of CLIL and immersion classrooms. *Language Learning Journal*, *42*(2), 181–194. https://doi.org/10.1080/09571736.2014.889509

Loewen, S. (2004). Uptake in Incidental Focus on Form in Meaning-Focused ESL Lessons. *Language Learning*, *54*(1), 153–188. https://doi.org/10.1111/j.1467-9922.2004.00251.x

Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, *27*(3), 361–386. https://doi.org/https://doi.org/10.1017/S0272263105050163

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, *19*(1), 1–14. https://doi.org/https://doi.org/10.1080/09588220600803311

Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *Modern Language Journal*, *90*(4), 536–556. https://doi.org/10.1111/j.1540-4781.2006.00465.x

Long, M. H. (1983). Does Second Language Instruction Make a Difference? A Review of Research. *TESOL Quarterly*, *17*(3), 359–382. https://doi.org/10.2307/3586253

Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Rowley, MA: Newbury House.

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & B. K. Bahtia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy. Essays in honor of A. Ronald Walton* (pp. 179–192). Philadelphia: John Benjamins.

Long, M. H. (2007). Recasts: The story thus far. In M. H. Long (Ed.), *Problems in SLA* (pp. 75–116). Mahwah, NJ: Lawrence Erlbaum.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, *26*(1), 27–56. https://doi.org/10.2307/3587368

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, *82*(3), 357–371. https://doi.org/10.1111/j.1540-4781.1998.tb01213.x

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in second language acquisition* (pp. 15–41). Cambridge, UK: Cambridge University Press.

Lowie, W., & Verspoor, M. (2015). Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation. *Language Learning*, *65*(1), 63–88. https://doi.org/10.1111/lang.12093

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*(3), 221–250. https://doi.org/https://doi.org/10.1177%2F136216880000400303

Lynch, T., & Maclean, J. (2001). A case of exercising: Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: second language learning, teaching, and testing* (pp. 141–162). Harlow, UK: Longman.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*(1), 51–81. https://doi.org/10.1017/S027226319800103X

Lyster, R. (2004). Differential Effects of Form-Focused Instruction. *Studies in Second Language Acquisition*, *26*, 399–432. https://doi.org/10.1017/S0272263104263021

Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, *59*(2), 453–498. https://doi.org/10.1111/j.1467-9922.2009.00512.x

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, *28*(2), 269–300. https://doi.org/https://doi.org/10.1017/S0272263106060128

Lyster, R., & Ranta, L. (1997). Corrective Feedback and Negotiation of Form in Communicative Classrooms. *Studies in Second Language Acquisition*, *19*(1), 37–66. https://doi.org/10.1017/S0272263197001034

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition*, *35*(1), 167–184. https://doi.org/10.1017/S027226311200071X

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, *32*(2), 265–302. https://doi.org/10.1017/S0272263109990520

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*(1), 1–40. https://doi.org/https://doi.org/10.1017/S0261444812000365

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, *60*(3), 501–533. https://doi.org/http://dx.doi.org/10.1111/j.1467-9922.2010.00565.x

Mackey, A., Gass, S. M., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 471–497. https://doi.org/10.1017/S0272263100004010

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–453). Oxford: Oxford University Press.

Mackey, A., Kanganas, A. P., & Oliver, R. (2007). Task Familiarity and Interactional

Feedback in Child ESL Classrooms. *TESOL Quarterly*, *41*(2), 285–312. https://doi.org/10.1002/j.1545-7249.2007.tb00060.x

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, *53*(1), 35–66. https://doi.org/10.1111/1467-9922.00210

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, *82*(3), 338–356. https://doi.org/10.1111/j.1540-4781.1998.tb01211.x

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–209). Philadelphia: John Benjamins.

Mackey, A., & Sachs, R. (2012). Older Learners in SLA Research: A First Look at Working Memory, Feedback, and L2 Development. *Language Learning*, *62*(3), 704–740. https://doi.org/10.1111/j.1467-9922.2011.00649.x

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, *13*(3–4), 422–429. https://doi.org/10.1080/09658210344000323

McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: A question of validity. *Journal of Speech and Hearing Disorders*, *47*, 19–24. https://doi.org/10.1044/jshd.4701.19

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 323 – 338). Oxford: Oxford University Press.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, *56*(4), 693–720. https://doi.org/10.1111/j.1467-9922.2006.00393.x

McDonough, K., & Nekrasova-Becker, T. (2014). Comparing the effect of skewed and balanced input on English as a foreign language learners' comprehension of the double-object dative construction. *Applied Psycholinguistics*, *35*(2), 419–442. https://doi.org/10.1017/S0142716412000446

Meara, P. (2005). *Llama language aptitude tests: The manual.* Swansea, UK: Lognostics.

Meara, P., Milton, J., & Lorenzo-Dus, N. (2003). *Swansea Language Aptitude Tests (LAT) v.2.0.* Lognostics.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*(1), 83–108. https://doi.org/10.1017/S0272263198001041

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL - International Review of Applied Linguistics in Language Teaching*, *45*(3), 241–259. https://doi.org/10.1515/iral.2007.011

Muranoi, H. (2000). Focus on Form Through Interaction Enhancement: Integrating Formal Instruction Into a Communicative Task in EFL Classrooms. *Language Learning*, *50*(4), 617–673. https://doi.org/10.1111/0023-8333.00142

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, *22*(1), 85–94. https://doi.org/10.3758/BF03202764

Nassaji, H. (2009). Effects of recasts and elicitations in dyadic interaction and the role of feedback explicitness. *Language Learning*, *59*(2), 411–452. https://doi.org/10.1111/j.1467-9922.2009.00511.x

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*(4), 719–758. https://doi.org/Doi 10.1111/0023-8333.00172

Nuevo, A. (2006). *Task complexity and interaction : L2 learning opportunities and development*. *Dissertation*. https://doi.org/10.13140/RG.2.1.5160.0807

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*(3), 377–402. https://doi.org/https://doi.org/10.1017/S0142716406060322

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological Memory Predicts Second Language Oral Fluency Gains in Adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/10.1017/S027226310707043X

Ohta, A. (2000). A learner-centered examination of corrective feedback in the Japanese classroom. In J. K. Hall & L. Verplaeste (Eds.), *The construction of second and foreign language learning through classroom interaction* (pp. 47–71). Mahwah, NJ: Lawrence Erlbaum.

Oliver, R., & Mackey, A. (2003). Interactional Context and Feedback in Child ESL Classrooms. *The Modern Language Journal*, *87*(4), 519–533. https://doi.org/10.1111/1540-4781.00205

Ortega, L. (1999). Planning and Focus on Form in L2 oral performance. *Studies in Second Language Acquisition*, *21*(1), 109–148. https://doi.org/10.1017/S0272263199001047

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pretask planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77–109). Philadelphia: John Benjamins.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, *36*(4), 573–595. https://doi.org/10.2307/3588241

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*(3), 331–347. https://doi.org/10.1016/0749-596X(91)90040-Q

Papagno, C., & Vallar, G. (1992). Phonological Short-term Memory and the Learning of Novel Words: The Effect of Phonological Similarity and Item Length. *The Quarterly Journal of Experimental Psychology Section A*, *44*(March 2014), 47–67. https://doi.org/10.1080/14640749208401283

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.

Payne, J. S., & Whitney, P. J. (2002). Synchronous CMC: Output, Working Memory, and Interlanguage Development. *Calico Journal*, *20*(1), 7–32. https://doi.org/10.11139/cj.20.1.7-32

Philp, J. (2003). Constraints on "noticing the gap": Non-native speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, *25*(1), 99–126. https://doi.org/https://doi.org/10.1017/S0272263103000044

Pienemann, M. (1984). Psychological Constraints on the Teachability of Languages. *Studies in Second Language Acquisition*, *6*(2), 186–214. https://doi.org/10.1017/S0272263100005015

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, *10*(1), 52–79. https://doi.org/10.1093/applin/10.1.52

Pienemann, M. (2015). An Outline of Processability Theory and Its Relationship to Other Approaches to SLA. *Language Learning*, *65*(1), 123–151. https://doi.org/10.1111/lang.12095

Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45–141). Adelaide, Australia: National Curriculum Research Centre, AMEP.

Pienemann, M., & Keßler, J. U. (2011). *Studying processability theory: An Introductory Textbook*. Amsterdam: John Benjamins.

Pinter, A. (2005). Task repetition with 10-year-old children. In C. Edwards & J. Willis (Eds.), *Teachers exploring tasks in English language teaching* (pp. 113–126). Basingstoke, UK: Palgrave Macmillan.

Pinter, A. (2007a). Some benefits of peer-peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research*, *11*(2), 189–207. https://doi.org/10.1177/1362168807074604

Pinter, A. (2007b). What children say: Benefits of task repetition. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 131–158). Newcastle, UK: Cambridge Scholars Publishing.

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626. https://doi.org/10.1111/lang.12010

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*(3), 437–470. https://doi.org/10.1017/S0272263109090366

Révész, A. (2012). Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures. *Language Learning*, *62*(1), 93–132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*, *38*(4), 703 – 737. https://doi.org/https://doi.org/10.1017/S0272263115000339

Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning*, *64*(3), 615–650. https://doi.org/10.1111/lang.12061

Révész, A., Sachs, R., & Mackey, A. (2011). Task complexity, uptake of recasts, and L2 development. In Second language task complexity. In P. Robinson (Ed.), *Researching the cognition hypothesis of language learning and performance.* (pp. 203–235). Amsterdam: John Benjamins.

Roberts, M. A. (1995). Awareness and the efficacy of error correction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning.* (pp. 163–182). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*, 283–331. https://doi.org/https://doi.org/10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language*

Learning, *45*(1), 99–140. https://doi.org/10.1111/j.1467-1770.1995.tb00964.x

Robinson, P. (1997). Individual Differences and the Fundamental Similarity of Implicit and Explicit Adult Second Language Learning. *Language Learning*, *47*(1), 45–99. https://doi.org/10.1111/0023-8333.21997002

Robinson, P. (2001a). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*(4), 368–392. https://doi.org/10.1177/026765830101700405

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2002). Learning conditions, aptitude complexes, and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113 – 133). Amsterdam: John Benjamins.

Robinson, P. (2003a). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Malden, MA: Blackwell.

Robinson, P. (2003b). The cognition hypothesis, task design, and adult task-based language learning. *Studies in Second Language Acquisition*, *21*(2), 45–105. https://doi.org/http://hdl.handle.net/10125/40656

Robinson, P. (2005). Cognitive Complexity and Task Sequencing: Studies in a Componential Framework for Second Language Task Design. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*(1), 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2007a). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 256–286). Cambridge Applied Linguistics.

Robinson, P. (2007b). Task complexity , theory of mind , and intentional reasoning : Effects on L2 speech production , interaction , uptake and perceptions of task difficulty, *45*, 193–213. https://doi.org/10.1515/IRAL.2007.009

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). Amsterdam: John Benjamins.

Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes, and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57–75). Springer, Berlin, Heidelberg.

Robinson, P., & Gilabert, R. (2007). Task complexity , the Cognition Hypothesis and, *45*, 161–176. https://doi.org/10.1515/IRAL.2007.007

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Philadelphia, PA: John Benjamins.

Sachs, R., & Suh, B. R. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 199–227). Oxford: Oxford University Press.

Sagarra, N. (2007a). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 229–248). Oxford: Oxford University Press.

Sagarra, N. (2007b). Working memory and L2 processing of redundant grammatical forms. In Z.-H. Han (Ed.), *Understanding second language process* (pp. 133–147). Clevedon: Multilingual Matters.

Sagarra, N., & Abbuhl, R. (2013). Optimizing the noticing of recasts via computer-delivered feedback: Evidence that oral input enhancement and working memory help second language learning. *Modern Language Journal*, *97*(1), 196–216. https://doi.org/10.1111/j.1540-4781.2013.01427.x

Saito, K. (2017). Effects of Sound, Vocabulary, and Grammar Learning Aptitude on Adult Second Language Speech Attainment in Foreign Language Classrooms. *Language Learning*, *67*(3), 665–693. https://doi.org/10.1111/lang.12244

Salaberry, M. R. (2000). L2 morphosyntactic development in text-based computer-mediated communication. *Computer Assisted Language Learning*, *13*(1), 5–27. https://doi.org/10.1076/0958-8221(200002)13

Sample, E., & Michel, M. (2014). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, *31*(8), 23–46. https://doi.org/https://doi.org/10.18806/tesl.v31i0.1185

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111–141). Philadelphia: John Benjamins.

Sato, M. (2011). Constitution of form-orientation: Contributions of context and explicit knowledge to learning from recasts. *Canadian Journal of Applied Linguistics*, *14*(1), 1–28. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=84358429&site=ehost-live

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, *13*(1), 96–120. https://doi.org/http://hdl.handle.net/10125/44170

Sauro, S. (2011). SCMC for SLA: A research synthesis. *Calico Journal*, *28*(2), 369–391. https://doi.org/10.11139/cj.28.2.369-391

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Service, E. (1992). Phonology, Working Memory, and Foreign-language Learning. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 21–50. https://doi.org/10.1080/14640749208401314

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, *32*(5), 608–623. https://doi.org/http://psycnet.apa.org/doi/10.1006/jmla.1993.1031

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16*(2), 155–172. https://doi.org/10.1017/S0142716400007062

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, *8*(3), 263–300. https://doi.org/10.1191/1362168804lr146oa

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, *4*, 361–392. https://doi.org/10.1191/1362168806lr203oa

Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 301–322). Oxford: Oxford University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning.* (pp. 69–93). Amsterdam: John Benjamins.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2014a). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*(3), 367–384. https://doi.org/https://doi.org/10.1093/applin/amu072

Skehan, P. (2014b). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). Amsterdam: John Benjamins.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition.* (pp. 17–40). Amsterdam: John Benjamins.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185–211. https://doi.org/10.1177/136216889700100302

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction.* (pp. 183–205). Cambridge, UK: Cambridge University Press.

Skehan, P., & Shum, S. (2014). Structure and processing condition in video-based narrative retelling. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 187–210). Amsterdam: John Benjamins.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York: Holt, Rinehart, and Winston.

Smith, B. (2004). Computer-mediated negociated interaction and lexical acquisition. *Studies in Second Language Acquisition*, *26*(3), 365–398. https://doi.org/doi:10.1017/S027226310426301X

Smith, B. (2010). Employing eye-tracking technology in researching the effectiveness of recasts in CMC. In F. M. Hult (Ed.), *Directions and prospects for educational linguistics* (pp. 79–98). New York: Springer.

Smith, B. (2012). Eye tracking as a measure of noticing: a study of explicit recasts in SCMC.

Language Learning & Technology, *16*(3), 53–81. https://doi.org/http://hdl.handle.net/10125/44300

Spada, N., Shiu, J. L. J., & Tomita, Y. (2015). Validating an Elicited Imitation Task as a Measure of Implicit Knowledge: Comparisons With Other Validation Studies. *Language Learning*, *65*(3), 723–751. https://doi.org/10.1111/lang.12129

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263–308. https://doi.org/doi10.1111/j.1467-9922.2010.00562.x

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*(2), 293–321. https://doi.org/10.1017/S0142716404001146

Suzuki, Y., & DeKeyser, R. (2015). Comparing Elicited Imitation and Word Monitoring as Measures of Implicit Knowledge. *Language Learning*, *65*(4), 860–895. https://doi.org/10.1111/lang.12138

Swain, M. (1991). French immersion and its offshoots: Getting two for one. In B. Freed (Ed.), *Foreign language acquisition: Research and the classroom* (pp. 91–103). Lexington, MA: Heath.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, *82*(3), 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Swan, M. (2005). *Practical English usage*. Oxford: Oxford University Press.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: John Benjamins.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183. https://doi.org/10.1017/S0272263100012870

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 171–195). Oxford: Oxford University Press.

Van De Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2015). Learning New Grammatical Structures in Task-Based Language Learning: The Effects of Recasts and Prompts. *Modern Language Journal*, *99*(2), 246–262. https://doi.org/10.1111/modl.12211

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Westport, CT: Ablex.

Vanpatten, B., Collopy, E., Price, J. E., Borst, S., & Qualin, A. (2013). Explicit Information, Grammatical Sensitivity, and the First-Noun Principle: A Cross-Linguistic Study in Processing Instruction. *Modern Language Journal*, *97*(2), 506–527. https://doi.org/10.1111/j.1540-4781.2013.12007.x

Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The Effects of Mode and Task Complexity on Second Language Production. *Language Learning*, *67*(2), 394–430. https://doi.org/10.1111/lang.12228

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, *25*(3), 315–339. https://doi.org/10.1093/applin/25.3.315

Wang, Z. (2014). L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 27–

62). Amsterdam: John Benjamins.

Wang, Z., & Skehan, P. (2014). Structure, lexis, and time perspective: Influences on task performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 155– 186). Amsterdam: John Benjamins.

Warschauer, M. (1996). Motivational aspects of using computers for writing and communication . In M. Warschauer (Ed.), *Telecollaboration in foreign language learning: Proceedings of the Hawai'i symposium* (pp. 29 – 46). Honolulu, HI : University of Hawai'i, Second Language Teaching and Curriculum Center.

Waters, G. S., & Caplan, D. (1996). The Measurement of Verbal Working Memory Capacity and Its Relation to Reading Comprehension. *The Quarterly Journal of Experimental Psychology A*, *49*(1), 51–79. https://doi.org/10.1080/027249896392801

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31. https://doi.org/https://doi.org/10.1017/S0261444816000276

Wickens, C. D. (2007). Attention to the second language. *IRAL - International Review of Applied Linguistics in Language Teaching*, *45*(3), 177–191. https://doi.org/10.1515/iral.2007.008

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*(1), 85–106. https://doi.org/10.1177/026553229701400105

Wilkins, D. A. (1976). *Notional syllabuses*. Oxford: Oxford University Press.

Williams, J. N., & Lovatt, P. (2003). Phonologiocal memory and rule learning. *Language Learning*, *53*(1), 67–121. https://doi.org/10.1111/1467-9922.00211

Yalçın, Ş., & Spada, N. (2016). Language Aptitude and Grammatical Difficulty. *Studies in Second Language Acquisition*, *38*(2), 239–263. https://doi.org/10.1017/S0272263115000509

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, *32*(2), 235–263. https://doi.org/10.1017/S0272263109990519

Yilmaz, Y. (2012). The Relative Effects of Explicit Correction and Recasts on Two Target Structures via Two Communication Modes. *Language Learning*, *62*(4), 1134–1169. https://doi.org/10.1111/j.1467-9922.2012.00726.x

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, *34*(3), 344–368. https://doi.org/10.1093/applin/ams044

Yilmaz, Y., & Granena, G. (2015). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*(1), 1–15. https://doi.org/10.1017/S136672891400090X

Yilmaz, Y., Granena, G., & Meyer, Z. S. (2016). The role of explicit language aptitude in implicit, explicit, and mixed feedback conditions. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition.* (pp. 327–349). Amsterdam: John Benjamins.

Yilmaz, Y., & Koylu, Y. (2016). The interaction between feedback exposure condition and phonetic coding ability. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition.* (pp. 303–326). Amsterdam: John Benjamins.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27. https://doi.org/https://doi.org/10.1093/applin/24.1.1

Zalbidea, J. (2017). “One Task Fits All”? The Roles of Task Complexity, Modality, and Working Memory Capacity in L2 Performance. *Modern Language Journal*, *101*(2), 335–352. https://doi.org/10.1111/modl.12389

Zhang, X., & Lantolf, J. P. (2015). Natural or Artificial: Is the Route of L2 Development Teachable? *Language Learning*, *65*(1), 152–180. https://doi.org/10.1111/lang.12094

Ziegler, N. (2015). Synchronous Computer-Mediated Communication and Interaction: A Meta-Analysis. *Studies in Second Language Acquisition*, *38*, 1–34. https://doi.org/10.1017/S027226311500025X

APPENDIX A

ΠΛΗΡΟΦΟΡΙΕΣ

Στοιχείο συμμετέχοντος (π.χ. S1, S2): ………..

Ηλικία: …………χρονών

Φύλο: κορίτσι □ αγόρι □

Τάξη: …………

Μητρική γλώσσα/γλώσσες: ………

Άλλες ξένες γλώσσες:………………………………………………………..

Διάρκεια εκμάθησης της Αγγλικής γλώσσας: ……… χρόνια

Παραμονή σε χώρα όπου ομιλείται η Αγγλική ως επίσημη γλώσσα : Ναι / Όχι

Εάν ναι, διάρκεια παραμονής σε χώρα όπου ομιλείται η Αγγλική ως επίσημη γλώσσα : ………μήνες ………..χρόνια

Πόσο συχνά γράφεις στον υπολογιστή; …………………

BACKGROUND INFORMATION

Participants (e.g. S1, S2): ………..

Age: …………years

Gender: girl □ boy □

Grade: …………

First: ………

Other Languages:………………………………………………………..

Length of Learning English: ……… years

Have you stayed in an English-speaking country? Yes / No

If yes, length of stay in an English-speaking country: ………months
………..years

How often do you type on a computer? …………………

APPENDIX B

DECISION-MAKING TASK 1

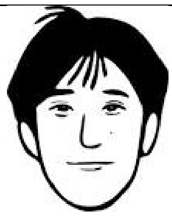
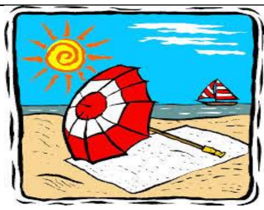
The characters in table 1 recently moved to a new block of flats. The transportation company left their items at the entrance to the building. You [the student] are the administrator of the building and you want Nektaria, who is your assistant, to return the items to their owners. Unfortunately, you don't have their phone numbers to ask them what they have lost. However, someone who knows them gave you information about what they usually do at the weekend and you are going to use this information from table 1 to find the owners of the items. Decide which items belong to whom and justify your answers to Nektaria by mentioning what they usually do at the weekend. Mention the time and whether they are at home so that Nektaria knows whether she will find them there.

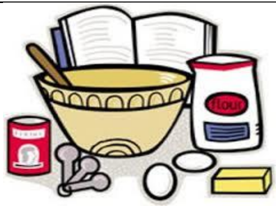
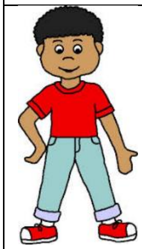
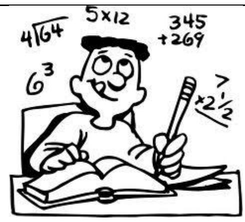
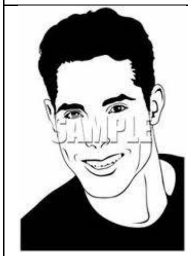
INFORMATION TRANSMISSION TASK 1

The characters in table 1 recently moved to a new block of flats. The transportation company left their items at the entrance to the building. Nektaria is the administrator of the building and she has been asked to return the items to their owners. Unfortunately, she doesn't have their phone numbers to ask them what they have lost. You [the student] know these people and what they usually do at the weekend. Nektaria needs this information to find their items. Give the information to Nektaria from table 1. Mention the activity, the time and whether they are at home so that Nektaria knows whether she will find them there.

TABLE 1 (It was the same for both the information transmission and decision-making tasks)

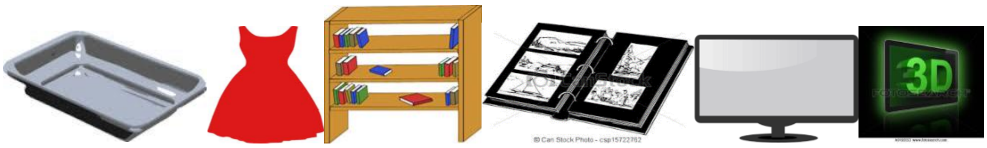
 HELEN	On Saturdays 9-11 PM At home	
 LUCY	On Saturdays 7-9 PM Not at home	

 MARY	On Sundays All day Not at home	
 MARK	On Saturdays 6-7 PM At home	
 JOHN	On Sundays 10-12 AM Not at home	
 JESSICA	On Saturdays 10-11 AM At home	
 NICK	On Sundays 4-6 PM At home	 
 JACK	On Saturdays 9-11 AM At home	

 NATALIE	On Sundays 3-5 PM At home	
 GEORGE	On Saturdays 9.00 AM At home	
 ALICE	On Sundays 10-12 AM At home	
 SOPHIE	On Sundays 11-12 AM At home	
 GREG	On Sundays 9-10 AM At home	
 TOM	On Saturdays 6-9 PM Not at home	

 BILL	On Saturdays 9-10 AM Not at home	

Items for decision-making task 1



DECISION-MAKING TASK 2

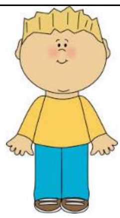
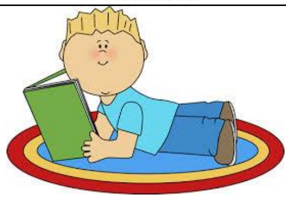
The characters in table 2 are tourists who lost some items at the airport. You [the student] work at the airport and you want Nektaria, who is your assistant, to return the items to their owners. However, the list of the owners of the items has been lost. Fortunately, you have information about what these people usually do at the weekend from a questionnaire they filled in. Use the information in table 2 and decide which items belong to whom. Justify your answers to Nektaria by mentioning what they usually do at the weekend. Mention the time and whether they are at home so that Nektaria knows whether she will find them there.

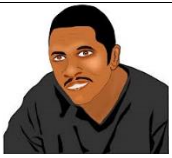
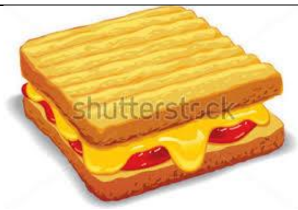
INFORMATION TRANSMISSION TASK 2

The characters in table 2 are tourists who lost some items at the airport. Nektaria works at the airport and she has been asked to return the items to their owners. You [the student] have information about what these people usually do at the weekend from a questionnaire they filled in. Nektaria needs this information to find the owners of the items. Give the information to Nektaria from table 2. Mention the activity, the time and whether they are at home so that Nektaria knows whether she will find them there.

TABLE 2 (It was the same for both the information transmission and decision-making tasks)

 CHRISTINE	On Saturdays 8-10 PM Not at home	
 ALEX	On Sundays 9-11 PM At home	
 HELEN	On Saturdays 7-8 PM Not at home	
 GEORGE	On Saturdays 8-10 PM At home	

 JACK	On Saturdays 5-6 PM Not at home	
 JOHN	On Sundays, 3-4 PM Not at home	
 MARK	On Sundays 6-7 PM At home	
 JAMES	On Saturdays 12-2 PM Not at home	
 SUSAN	On Saturdays 7-9 PM Not at home	
 NICK	On Saturdays 9-10 AM Not at home	
 JANE	On Sundays 1-2 PM At home	

 ALAN	On Sundays All day Not at home	
 BILL	On Saturdays 10-12 AM At home	
 KATE	On Sundays 11.00 AM At home	
 CHRIS	On Sundays 10-11 AM At home	

Items for decision-making task 2












APPENDIX C

ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΑΣΚΗΣΗ

Παρακαλώ κάνε τικ στο κουτί που αντιστοιχεί στην απάντησή σου στην παρακάτω κλίμακα.

1.	Η άσκηση δεν απαιτούσε καθόλου πνευματική προσπάθεια.	☐	☐	☐	☐	☐	☐	☐	☐	☐	Η άσκηση απαιτούσε πολλή πνευματική προσπάθεια.
2.	Η άσκηση δεν ήταν καθόλου δύσκολη.	☐	☐	☐	☐	☐	☐	☐	☐	☐	Η άσκηση ήταν υπερβολικά δύσκολη.
3.	Δεν είχα καθόλου πρόβλημα με την χρήση της γλώσσας στη διάρκεια της άσκησης.	☐	☐	☐	☐	☐	☐	☐	☐	☐	Είχα πολλά προβλήματα με τη χρήση της γλώσσας στη διάρκεια της άσκησης .
4.	Τα πήγα πολύ καλά στην άσκηση.	☐	☐	☐	☐	☐	☐	☐	☐	☐	Δεν τα πήγα καθόλου καλά στην άσκηση.

QUESTIONS ABOUT THE TASK

Please tick the table that corresponds to your answer in the following scale.

1.	The task required no mental effort.	☐	☐	☐	☐	☐	☐	☐	☐	☐	The task required a lot of mental effort.
2.	The task was not difficult at all.	☐	☐	☐	☐	☐	☐	☐	☐	☐	The task was very difficult.
3.	I had no problems while using English during the task.	☐	☐	☐	☐	☐	☐	☐	☐	☐	I had a lot of problems while using English during the task.
4.	I did very well.	☐	☐	☐	☐	☐	☐	☐	☐	☐	I didn't do well.

APPENDIX D

ORAL PRODUCTION PRE-TEST

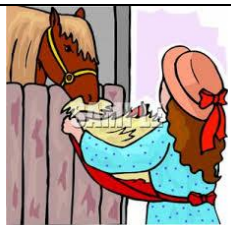
The table below shows how the characters in the pictures spend their time at the weekends. Give this information to Nektaria orally.

Tom Saturday 9.00 a.m.	
Nick Saturday 10.00 a.m.	
Helen Sunday 11.00 a.m.	
Mark Sunday 9.00 a.m.	
John Saturday 4.00 p.m.	
Sophie Sunday 10.00 a.m.	
Andrew Saturday 8.00 a.m.	

Sam Saturday 10.00 a.m.	
Laura Saturday 11.00 a.m.	
Craig Sunday 5.00 p.m.	
Jack Sunday 9.00 a.m.	
Chris Sunday 11.00 a.m.	

ORAL PRODUCTION POST-TEST

The table below shows how the characters in the pictures spend their time at the weekends. Give this information to Nektaria orally.

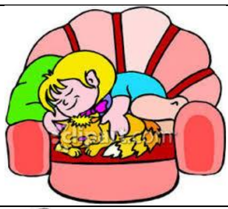
Tom Saturday 9.00 a.m.	
Nick Saturday 10.00 a.m.	
Helen Sunday 11.00 a.m.	
Mark Sunday 9.00 a.m.	
John Saturday 4.00 p.m.	
Sophie Sunday 10.00 a.m.	
Andrew Saturday 8.00 a.m.	

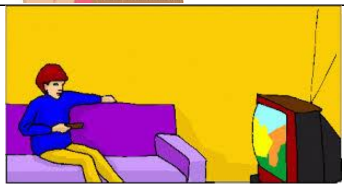
Sam Saturday 10.00 a.m.	 
Laura Saturday 11.00 a.m.	
Craig Sunday 5.00 p.m.	
Jennifer Sunday 9.00 a.m.	
Georgia Sunday 11.00 a.m.	

APPENDIX E

WRITTEN PRODUCTION PRE-TEST

The table below shows how the characters in the pictures spend their time at the weekends. Write this information to Nektaria.

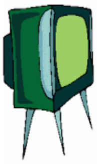
Georgia Saturday 9.00 a.m.	
Jennifer Saturday 10.00 a.m.	
Nick Sunday 11.00 a.m.	
Mark Sunday 9.00 a.m.	
John Saturday 4.00 p.m.	
Jack Sunday 10.00 a.m.	

Andrew Saturday 8.00 a.m.	
Sam Saturday 10.00 a.m.	
Laura Saturday 11.00 a.m.	
Craig Sunday 5.00 p.m.	
Sophie Sunday 9.00 a.m.	
Christine Sunday 11.00 a.m.	

WRITTEN PRODUCTION POST-TEST

The table below shows how the characters in the pictures spend their time at the weekends. Write this information to Nektaria.

Kate Saturday 9.00 a.m.	
Nick Saturday 10.00 a.m.	
Helen Sunday 11.00 a.m.	
Mark Sunday 9.00 a.m.	
John Saturday 4.00 p.m.	
George Sunday 10.00 a.m.	
Andrew Saturday 8.00 a.m.	

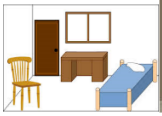
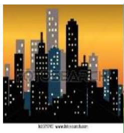
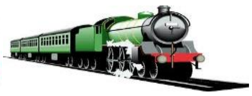
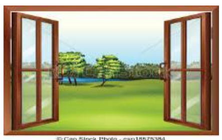
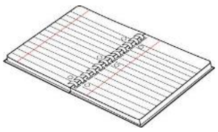
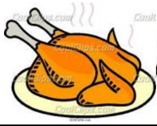
Sam Saturday 10.00 a.m.	  
Laura Saturday 11.00 a.m.	
Craig Sunday 5.00 p.m.	 
Jack Sunday 9.00 a.m.	
Chris Sunday 11.00 a.m.	

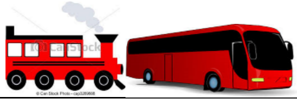
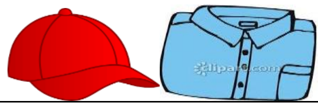
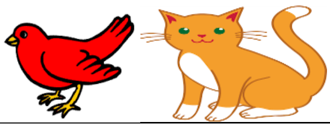
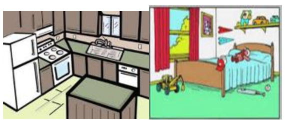
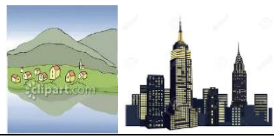
APPENDIX F

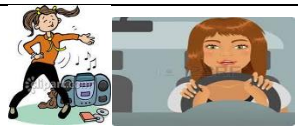
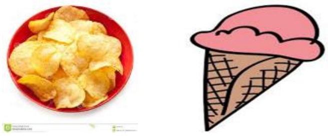
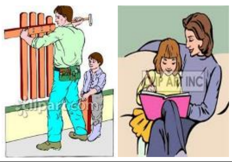
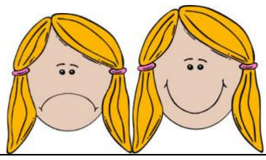
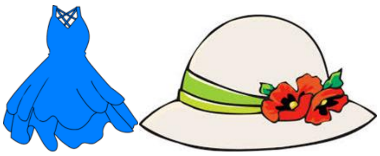
You will listen to some sentences only once. After you listen to the sentence, do the following:

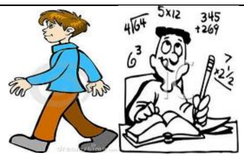
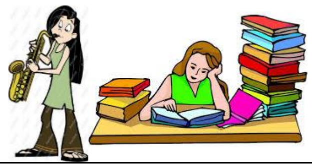
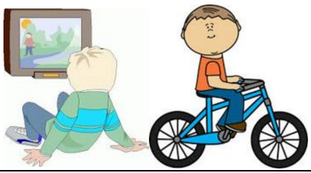
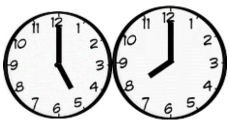
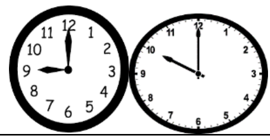
-Choose the correct picture. If the correct picture is A, say A. If the correct picture is B, say B (*the pictures had the letters A and B on the slides*).

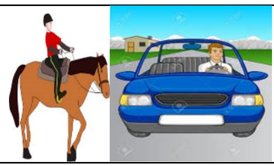
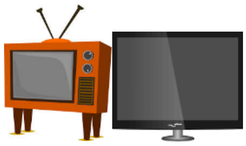
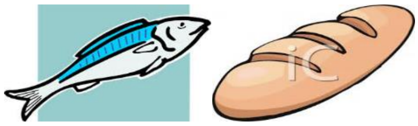
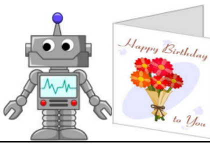
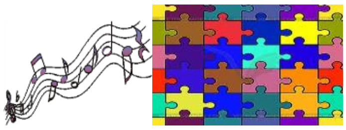
-After you choose the correct picture, wait for the colour to change. Then, repeat the sentence in correct English.

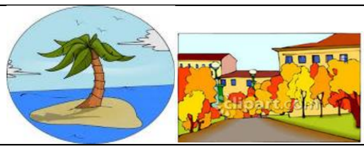
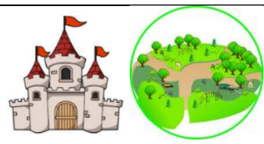
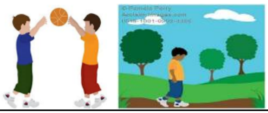
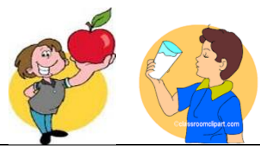
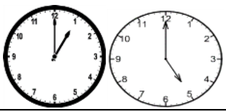
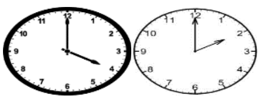
EI PRETEST	EI POSTTEST
I'm reading a book now.  	I'm writing a book now.  
The kitchen was such small.  	The village was such small.  
He always take the bus.  	He always take the train.  
I'm open the door now.  	I'm open the bag now.  
She's wearing a nice hat.  	She's wearing a nice skirt.  
He sometimes gives a sweet.  	He sometimes gives a cake.  
My brother made this cake.	My brother made this film.

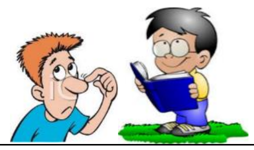
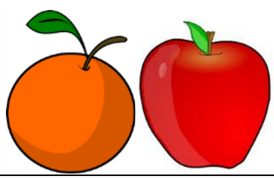
	
She’ll cooking for her friend. 	She’ll cooking for her dad. 
He often miss the bus. 	He often miss the train. 
I wanted wear a dress. 	I wanted wear a shirt. 
Her salad was the best. 	Her salad was the worst. 
He always sleeps at home. 	He always sleeps at work. 
My sister found a dog. 	My sister found a cat. 
The kitchen was so small. 	The village was so small. 
He often cry at night. 	He often cry at home. 
My sister can to drive.	My sister can to sing.

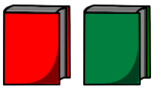
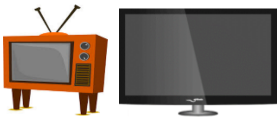
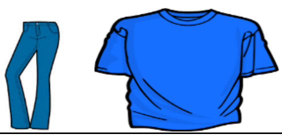
	
He arrived in France late. 	He arrived in Greece late. 
He always uses a spoon. 	He always uses a glass. 
I'm going to buy chips. 	I'm going to buy sweets. 
She opened big brown bag. 	She opened big black bag. 
He sometimes help his dad. 	He sometimes help his mum. 
I'm going in a town. 	I'm going in a pub. 
My brother was so sad. 	My sister was so sad. 
She often buys a dress. 	She often buys a hat. 
I studied Maths at night.	I studied Greek at night.

	
The city was much big. 	The sofa was much big. 
He sometimes watch a film. 	He sometimes watch a match. 
My father left in eight. 	My father left in ten. 
Our office was so clean. 	Our kitchen was so clean. 
She always makes a cake. 	She always makes a card. 
The baby must eat now. 	The children must eat now. 
He studied Maths on home. 	He studied Maths on school. 
He sometimes drive a car.	He sometimes drive a bus.

	
My mother bought red hat. 	My mother bought red dress. 
Your mobile was so old. 	Your TV was so old. 
He sometimes teaches a dance. 	He sometimes teaches a sport. 
I needed to buy tea. 	I needed to buy bread. 
My brother found mine hat. 	My brother found mine car. 
She always work at ten. 	She always work at home. 
My brother did this card. 	My brother did this song. 
She’ll visit a small town.	She’ll visit a small park.

	
He sometimes plays with George. 	He sometimes plays with Nick. 
My mother bought a car. 	My mother bought a bag. 
Her salad was much worst. 	Her salad was much best. 
He always wash his hands. 	He always wash his car. 
I needed buy some cheese. 	I needed buy some milk. 
The sofa was too big. 	The city was too big. 
He sometimes drinks the juice. 	He sometimes drinks the milk. 
My father left at five. 	My father left at two. 
He studied Maths at school. 	He studied Maths at home. 

She often need a pen. 	She often need a chair. 
The baby must to sleep. 	The children must to sleep. 
She opened a black bag. 	She opened a brown bag. 
He sometimes touches his head. 	He sometimes touches his nose. 
My sister can swim well. 	My sister can cook well. 
He arrived to Spain late. 	He arrived to Rome late. 
He often speak with Nick. 	He often speak with George. 
I'm going buy a dress. 	I'm going buy a hat. 
My brother found my bag. 	My brother found my apple. 
He sometimes stays at home.	He sometimes stays at work.

	
I'm going to a beach. 	I'm going to a park. 
My mother was such sad. 	My father was such sad. 
She always dance with Nick. 	She always dance with George. 
My brother found green book. 	My brother found green pen. 
Your TV was such old. 	Your mobile was such old. 
He often breaks a plate. 	He often breaks a glass. 
I wanted to wear boots. 	I wanted to wear jeans. 
She's travel with her friend. 	She's travel with her mum. 
He sometimes call his mum.	He sometimes call his dad.

	
I studied French in night. 	I studied hard in night. 
Ours kitchen was so clean. 	Ours office was so clean. 
He always closes the door. 	He always closes the book.